ESSAYS ON TERM STRUCTURE, FORWARD PREMIUM ANOMALY AND GLOBALIZATION

by

Ting Ting Huang

MA in Economics, Youngstown State University, 2003

Submitted to the Graduate Faculty of

the the School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Ting Ting Huang

It was defended on

May 13th 2010

and approved by

Steven Husted, Dept. Economics

Xinfu Chen, Dept. Mathematics

James Cassing, Dept. Economics

James Feigenbaum, Dept. Economics

Dissertation Advisors: Steven Husted, Dept. Economics,

Xinfu Chen, Dept. Mathematics

ESSAYS ON TERM STRUCTURE, FORWARD PREMIUM ANOMALY AND GLOBALIZATION

Ting Ting Huang, PhD

University of Pittsburgh, 2010

Term structure models have attracted tremendous amount of attention in the last two decades. My first paper specifies the dynamic and cross-sectional behavior of bonds in the framework of the Linear or general affine term structure model (ATSM). After revisiting the basic theory of ATSM under the physical probability measure, a linear version, LTSM is proposed. We find theoretical loads of ATSM and LTSM by solving Riccati equations, with parameters chosen for the solution to match that from the principal two–component models. This paper is the first which provides an empirical model, and numerically studies the state spaces that guarantee the Black–Scholes equation is uniquely solvable and that the yields are always positive, so it clarifies the condition of Duffie and Kan (1996) [42].

The second paper utilizes the LTSM to study the forward premium anomaly. This allows me to model the behavior of the risk premium theoretically and empirically. I test my model using data on the Canadian-U.S. exchange rate. The dynamic factors are captured by Composite Principal Component Analysis (CPCA) which supplies a different way to set up the global factors for both currencies. Different from previous work in this area, the LTSM and ATSM can account for the anomaly, and the theoretical interest rates are guaranteed to be positive.

Freund and Weinhold (2000, 2002 & 2004) stated that the Internet stimulates

international trade. However, the data provided by World Bank is certainly not an ideal measure. My third paper presents a new way to measure the diffusion of the Internet, downloaded from Cooperative Association for Internet Data Analysis (CAIDA). The major findings are that a significant and positive relationship exists between Internet distance and the bilateral international trade volumes across ten countries. Furthermore, the magnitude of elasticity is discussed and further support the conclusion of Freund and Weinhold.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I am grateful to my advisor Professor Xinfu Chen, not only for his guidance and advice throughout the completion of my thesis but also for giving me four wonderful years of researching and teaching at the University of Pittsburgh. I also want to thank to my other advisors, Professor Steven Husted, Professor James Cassing and Professor James Feigenbaum for joining the committee for my defense and overview. Many thanks go to my dear husband Bruce Qiang Sun for fruitful and encouraging discussions. I am also grateful to my department colleagues for many valuable comments on my research and for sharing their knowledge in finance and economic theory. Finally, I appreciated the support and encouragement from my family throughout the entire four years of working on these papers.

1.0 INTRODUCTION

My dissertation is composed of three essays that deal with a variety of topics in financial economics, international finance, and empirical international trade. The main contributions of this dissertation include showing the theoretical and empirical explanation to the dynamic and cross-sectional behavior of bonds in the framework of the Linear and Affine Term Structure Models together with how the common factors are applied in financial market problems and the development and use of a new data set on Internet usage in order to see the impact of the internet on international trade.

The first essay is titled “A Theoretical and Empirical Analysis of Linear Term Structure Models”. Term structure models deal with the dynamic and cross-section behavior of interest rates of various maturities. Linear and affine versions of these models relate the behavior of interest rates to underlying common yield factors. Most previous work on these models has assumed that actual interest rates are markups over an unobservable risk neutral rate of return. Under this assumption and an assumption about the statistical distribution of this rate of return, estimates of the term structure are generated via simulation techniques. In this chapter, I use actual data on the actual distribution of interest rates and then employ principal components analysis to estimate the common factors. Given these, I then construct a general affine term structure model. This work represents an extension of work I published in the monograph Theoretical and Empirical Analysis of Common Factors in a Term Structure Model (Cambridge Scholars Publishing, 2009).

I compute empirical common factors for ten U.S. government bonds using daily data for the period 1993-2006. I find that two common factors are sufficient to account for 99 percent of the total variance of interest rates. To test for their independence, I calculate the empirical copula (the joint distribution of the transformed random variables by their marginal distribution functions) of the common factors. Finally, I show that this version of the Linear Term Structure Model is remarkably successful at capturing the behavior of the yield curve across time and maturities.

My second essay, titled "Linear Term Structure Models and the Forward Premium Anomaly", uses the techniques developed in the first essay and applies them to well known puzzle in the international finance literature, the forward premium anomaly. This anomaly refers to the fact that many empirical studies of exchange rate behavior appear to show that the forward exchange rate, which is determined by international interest rate differentials, is a biased predictor of the expected future spot rate. This should not be the case if a commonly assumed foreign exchange market theoretical equilibrium condition known as Uncovered Interest Rate Parity (UIP) were to hold. Fama (1984) argues this bias may be due to the presence of an unobservable risk premium, and then shows how the risk premium must behave in order to be consistent with empirical studies. In particular, he argues that the variance of the risk premium must be larger than the variance of the expected rate of change in the exchange rate and that the covariance between these two series must be negative.

Backus, Foresi, and Telmer (2001) use the affine term structure model (ATSM) to study the anomaly. However, under the assumption that the global factors follow a special distribution, they cannot rule out that the theoretical interest rate can be negative with positive probability. Following the lead from the earlier paper, I rely on actual data rather than relying on assumed behavior, I set out to develop estimates of the theoretical risk premium.

I estimate my model using data on the Canadian-U.S. exchange rate. The dynamic factors are captured by the Composite Principal Component Analysis (CPCA). In this paper, the expected excess returns are represented by the risk premium associated in the risk-adjusted UIP relationship. Using these estimates, I provide some evidence on the interrelationship between the expected rate of depreciation and the forward premium; the fluctuation of the risk premium is shown to be greater than the forward premium. My results are clearly consistent with the Fama's conditions.

The third essay, titled "An Investigation of New Internet Measurement On International Trade", is about the Internet measurement, international trade and computational data mining. This paper seeks to answer the question: What kind of role does the Internet play in international trade? So far, because of a lack of strong evidence and data, there is a huge difference in the point of view about this question. It is often argued that influences of the diffusion of the Internet on international trade are almost everywhere. But testing this proposition is hindered by a significant constraint: a shortage of the right data.

A principal focus of this paper is the development of a new set of data on Internet usage. The data measure Internet cross-traffic by quantifying several attributes, including the round trip times, which characterize macroscopic connectivity and performance of the Internet and allow various topological and geographical representations at multiple levels of aggregation granularity. I then use this data to re-investigate earlier studies of the relationship between trade and the Internet usage. In so doing, I confirm the findings of Freund and Weinhold (2000) that the Internet usage is significantly related to trade.

2.0 A THEORETICAL AND EMPIRICAL ANALYSIS OF LINEAR TERM STRUCTURE MODELS

2.1 INTRODUCTION

Term structure refers to the dynamic and cross-sectional behavior of bonds, and term structure models have been playing the central roles in today's financial modeling. Since the pioneer work of Vasicek [90] (1977), there has been a significant amount of progress towards the term structure model; see, for instance, Cox, Ingersoll, and Ross [31] (1985), Ho and Lee [58] (1986), Black, Derman, and Troy [18] (1990), Heath, Jarrow and Morton [57](1992), Duffie and Kan [42] (1996), Duffie, Pan and Singleton [43] (2000), Wagner [66] (2006) and the references therein.

While theoretically well–studied, the ATSM was empirically studied by Piazzesi [81] (2003), partially with the help of principal component analysis(PCA) [1]. Our paper extends the work of Piazzesi [81] to the full extent of ATSM. Following Piazzesi, we shall use principal components as state variables to construct an affine term structure model, and linear term structure model(LTSM), in which formulas are valid under the physical probability measure. As we use 3313 sets of empirical daily data which are computed to get the zero-coupon bonds with maturities $1/4, 1/2, 1, 2, 3, 5, 7, 10, 20, 30$ (years) from 1993 to 2007, we find PCA to get loads, $l_i(.)$, and a sample path of the

[1]Most literature focuses on the estimation parts for two factors or three factors in different interesting applications. The number of factors was not specified, unless based on each specific issue. They mostly show the results for both examinations, and some show three–factors case following Litterman and Scheinkman [71] (1991).

factor, X_t^i. Both principal two–component models capture more than 99 percent of the total variances and more than 99.9 percent of total square of the norms. Thus, basically all information about the yields are characterized by the two statistical factors and the corresponding loads. While this part is similar to that of Piazzesi [81], we go further finding stochastic models for the factors. According to the ATSM theory of Duffie and Kan [42] (1996), the factors abide by the stochastic differential equations where the random process is a standard n-dimensional Winner process whose martingale-measure is the risk-neutral measure. In empirical estimation of parameters, it may not be compatible to start with this risk-neutral settings. In our setting, there is no restriction on the affine structure on the drift term. This gives us extra room in choosing the state space of factors.

This paper has the following new ingredients. I. We develop a new method to determine the parameters and the empirical loads for the Riccati equation. We firstly evaluate the regression of the variance to the factors, by starting from the initial approximation, then running the Newton's Iteration to solve a minimization problem. So we got a set optimal parameters in the Riccati equations. The parameters we obtained provide the short-rate and prices of risks, thereby completely determining the famous Black-Scholes partial differential equation [20, 76]. Our numerics show that theoretical loads match empirical ones from principal component analysis quite well. The resulting LTSM and ATSM present yield surfaces that fit the empirical one at all points.

II. Our model is built upon the physical probability measure. Indeed, the artificial risk neutral probability measure can be omitted from the theorems. As a result, any data from reality can be used as i.i.d random samples obeying the model assumptions, then can be analyzed with canonical econometrics methods like the PCA argument. This is important since the samples may not be regarded as i.i.d random variables under the risk neutral measure with the separated distributions, but risk neutral probability measure is artificial. For example, in Black-Schole theory, $dS_t = \mu dt + \sigma dW_t$, the natural observations can not be used to determine parameters in the risk–neutral equation $dS_t = rdt + \sigma dW_t^*$. At least, corrections are needed here. Based on physical probability measure throughout the whole theorem framework, instead of the artificial risk–neutral measure, we establish our theorems, which is clearly different from the previous literature. This can supply the solid background to have our model fit well the empirical data. For the general setting of term structure, our theorem does not have to begin with the assumption that the short–term rate of return is risk–free.

III. We reviewed the ATSM in a concrete way and supply a new model, LTSM. In our LTSM, the number of parameters need to evaluated is 12 where as ATSM need 18 of them. The proposed linear term structure models greatly drop the total number of parameters needed to be estimated in the Riccati equation, which makes the calibration feasible and even efficient, besides capturing the time–varying properties of expected mean rate of bonds returns to different maturities.

IV. This paper shall provide the models with factors state space that guarantees that the yields are always positive and this is missing in most literature. The state space we used guarantees the Black-Scholes equation admits a unique solution.

Here are the general reviews of the related research works in affine term structure models. In term structure models, yields of bonds are modeled by state variables, such as yield factors, common factors which are mainly macroeconomic fundamentals, or principal components which are statistical factors. Usually, there are problems in term structure models as to how to draw the information at any history of time while a huge number of nominal bonds are traded. With the help of common factors, one can, with as little loss of information as possible, reduce large-dimensional data to a limited number of factors. Litterman and Scheinkman [71] (1991) used a Principal Component Analysis (PCA) on three US treasury bonds to estimate common factors. Pagan, Hall, and Martin [78] (1995) used stylized factors that pertain to the nature for the term structure modeling. Baum and Bekdache [16] (1996) applied stylized factors to the dynamics of short, medium, and long-term interest rates and explained the factors by incorporating asymmetric GARCH representations. Connor [30] (1995) and Campbell, Lo, and MacKinlay [22] (1997) characterized three types of factor models: the known-factor model, the fundamental-factor model, and the statistical-factor model. The statistical factors, as discussed by Alexander [4] (2001) and Zivot and Wang [93] (2003)), can be modeled by principal components. Cochrane [28] (2001) summarized that the pricing kernel was linear in the factors both in the economic time series and in the pricing models. In this paper we shall provide a new point of view of using principal components to investigate the yield-to-maturity curves.

In 1996, Duffie and Kan [42] systematically studied a special class of term structures, the Affine Term Structure Model (ATSM); here the term “affine” refers to the assumption that yields are affine functions of factors, or state variables. The coefficients of the factors, called loads, are solutions of ordinary differential equations of Riccati type. Using yields themselves as factors, they provided a few simulated numerical examples, in which yields follow a parametric multivariate Markov diffusion process with standard Brownian Motions. They provided conditions on the stochastic differential equations for this affine representations under the risk–neutral measure throughout the whole paper, which is artificial and non-observable. Piazzesi [80] (1998) considered an ATSM with jumps in several macroeconomics frameworks. Dai and Singleton [32] (2000) studied some econometrics issues and autoregressive structural differences, and pointed out the trade–off flexibility in choosing between the conditional correlations and the volatilities of the risk factors. Duffie, Pan and Singleton [43] (2000) greatly extended the framework of ATSM to a wide range of valuation and econometric problems, to defaultable corporate bonds, and to include jumps, while in principle using the yields as factors. Dai and Singleton [33] (2003) posted a critical survey of four different term structure models, in order to check whether the theoretical specification of term structure models matches the yield curves. They reviewed term structure models under the risk–neutral measure, and checked the fitting of the models by matching linear coefficients for changes in yields with the slope of the yield curve, and the possibility of producing hump-shaped unconditional yield volatilities. Then they drew conclusions that their overview of the empirical fit of dynamic term structure models (DTSMs) had underscored several successes, while highlighting several challenges for future research. Further, Ahn, Dittmar found that their most flexible model had less volatility than the observed historical data. It seems that at maximum-likelihood estimates of the parameters, a tractable term structure model is still undiscovered. In [34] (2006), they developed discrete-time, nonlinear term structure models. Under the risk-neutral measure, the discrete-time affine processes are the counterparts of models in Duffie and Kan (1996) and Dai and Singleton (2000). They use the market price of risk to link the risk-neutral and historical distributions on the state variables, with the closed form of conditional likelihood functions for coupon bond yields. Their results show that inclusion of a cubic term in the drift significantly improves the models statistical fit as well as its out-of-sample forecasting performance.

Affine and linear term structure models can be applied in many economics and finance issues. Here are the basic reviews of the applications. Diebold, Piazzesi and Rudebusch [37] (2005) comprehensively illustrated the importance of understanding what moves bond yields and usefulness of factor models, how should macroeconomic variables be combined with yield factors, and what are the links between macro variables and yield–curve factors. One of the research objectives is to do the derivative pricing and hedging. The price of security derivatives, such as swaps, caps and floors, futures and options on interest rates are computed from the given model of the yield-curve [43]. Banks need to manage the financial and credit risk on loans. For those contracts that are contingent on future short rates, such as swap contracts. brokers need to apply appropriate hedging strategies to master the uncertainty of the economy. Second, consumption-based asset pricing models are also under consideration. Davis and Heathcote (2005) explored general equilibrium models with housing, which showed the implications of a real business cycle model with a construction sector. Ortalo-Magne and Rady (2006) analyzed an overlapping generations model to study prices and volume in the housing market. Cocco (2005), Flavin and Yamashita (2002), and Flavin and Nakagawa (2005) considered portfolio choice with exogenous returns in the presence of housing. Piazzesi, Schneider and Tuzel (2007) derived the effects of housing on asset prices in a general equilibrium model. Piazzesi and Schneider (2007) considered the role of inflation affecting the pricing of nominal bonds. They put the analysis framework in a representative agent asset pricing model with recursive utility preferences and exogenous consumption growth and inflation. They also coauthored in another paper to discuss the asset pricing in a general equilibrium model in which some agents suffer from inflation illusion. Third, the Fed needs to control the monetary policies and to construct debt plans in the macro–equilibrium environment. This is another reason for studying the yield curve. The expectations hypothesis suggested by Balduzzi, Bertola, and Foresi (1996) shows how the transmission mechanism works. Piazzesi (2005) discovered bond yields respond to policy decisions by the Federal Reserve and vice versa, by modeling a high-frequency policy rule based on yield curve under arbitrage-free.

The rest of the paper is organized as follows: Section 2 reviews ATSM and Black-Scholes theory. Section 3 provides a further discussion on the Black-Scholes equation in ATSM framework. Section 4 describes theoretical and numerical derivation of PCA that we shall use. Section 5 presents our empirical work: first we perform PCA to obtain empirical factors and loads needed in ATSM and LTSM; then we perform linear regression to obtain covariance matrix of the innovation of factors; finally we solve the Riccati equations to obtain theoretical loads for ATSM and LTSM. In Section 6 we tune the empirical LTSM and ATSM so that they are equipped with good state spaces. Section 7 concludes the paper.

2.2 TERM STRUCTURE

2.2.1 Term Structure

A term structure models time t value of T-bond for any $t \geqslant 0$ and $T > t$. Here by **T-bond** it means *a guaranteed payment of unit amount at time T*. For $t < T$, we use Z_t^T to denote the price of one share of T-bond at time t. By default, $Z_T^T = 1$. Following benchmark models (e.g. [57, 42]), we consider the assumption

(A1) *the collection $\{Z_t^T\}_{0 \leqslant t \leqslant T, T>0}$ obeys a stochastic differential equation*

$$\frac{dZ_t^T}{Z_t^T} = \mu_t^T \; dt + \sum_{i=1}^{n} \sigma_{t\,i}^T \;\; dX_t^i \tag{2.2.1}$$

where $dZ_t^T = Z_{t+dt}^T - Z_t^T$, $\{(X_t^1, \cdots, X_t^n)\}$ is a stochastic process and $\{\mu_t^T, \sigma_{t1}^T, \cdots, \sigma_{tn}^T\}$ are stochastic processes adapted to a natural filtration[2].

A system is **arbitrage-free** if *the possibility that one can make guaranteed profit out of nothing is zero*. A fundamental theory on Term Structure Model (TSM) is the following:

Theorem 1 (Theory of Term Structure Model)**.** *Assume* **(A1)** *in an arbitrage-free system. Then there exist processes $\{R_t, P_t^1, \cdots, P_t^n\}$ adapted to a natural filtration such that*

$$\mu_t^T = R_t + \sum_{i=1}^{n} P_t^i \, \sigma_{t\,i}^T \qquad \forall\, T > 0, t \in [0, T]. \tag{2.2.2}$$

The proof will be given in the Appendix A.1.

Remark 2.2.1. (1) We prove Theorem 1 *without* the We ignore the traditional assumption that short-term bonds are risk-free and as functions of T, $\sigma_{t1}^T, \cdots, \sigma_{tn}^T$ are linearly independent for some t.

---

[2] A process $\{x_t\}_{t \in \mathbf{T}}$ is **adapted to a natural filtration** if x_τ is observable at any time $t \geqslant \tau \in \mathbf{T}$.

Here by short-term bonds being risk-free it means that buying a $(t+dt)$-bond at time t and selling it at time $t+dt$ produces a fixed return rate, R_t, called **short-rate**. This assumption can be implemented (e.g. [57]) by assuming $\sigma_{t\,i}^{t+dt} = 0$ for all i, t. Then (2.2.2) gives $R_t = \mu_t^{t+dt}$ and (2.2.1) implies $Z_t^{t+dt} = e^{-R_t dt}$.

(3) Typically $\{(X_t^1, \cdots, X_t^n)\}$ in (2.2.1) is assumed to be a martingale under a measure $\mathbb{P}$ of natural observation, so μ_t^T is the observed expected return rate of the *investment on T-bond:* purchasing a T-bond at time t and selling it at time $t+dt$. The identity (2.2.2) proclaims that any increment of the expected return from the short-rate R_t can only be achieved with risks. In (2.2.2), the multiple P_t^i of the volatility $\sigma_{t\,i}^T$ is therefore called the **price of risk** on the **uncertainty innovation** $\sigma_{ti}^T dX_t^i$.

(4) By (2.2.2), the TSM (2.2.1) can be written as

$$\frac{dZ_t^T}{Z_t^T} = R_t\, dt + \sum_{i=1}^n \sigma_{t\,i}^T\, dX_t^{*i}, \qquad X_t^{*i} := X_t^i + \int_0^t P_s^i ds \qquad \forall\, i = 1, \cdots, n,\ t \geqslant 0.$$

Suppose $\mathbb{Q}$ is a measure under which $\{\{X_t^{*i}\}_{t\geqslant 0}\}_{i=1}^n$ are martingales. Then under $\mathbb{Q}$, the expected rate of return of the investment on T-bond is R_t, for any $T > t$. This particular measure $\mathbb{Q}$ is called the **risk-neutral measure**.

(5) Under certain non-degeneracy assumption on the uncertainty innovation $\{\{dX_t^i\}_{t\geqslant 0}\}_{i=1}^n$, (2.2.2) is indeed a necessary and sufficient condition for (2.2.1) to be arbitrage-free.

2.2.2 Affine Term Structure

As a special case of TSM, the affine term structure model, **ATSM**, assumes that the logarithms of bond prices are affine functions of factors; i.e,

(A2) *the price Z_t^T of the T-bond at time t satisfies*

$$\log \frac{1}{Z_t^T} = A_0(T-t) + \sum_{i=1}^n A_i(T-t)\ X_t^i \qquad \forall\, t \geqslant 0, T \in [t, t+T^{\max}) \qquad (2.2.3)$$

where $A_0(\cdot), \cdots, A_n(\cdot)$ are differentiable functions defined on $[0, T^{\max})$ and $\{(X_t^1, \cdots, X_t^n)\}$

is an Itô process with a positive definite covariance matrix $\mathrm{Cov}(X_\tau^1, \cdots, X_\tau^n)$ *for some* $\tau \geqslant 0$. We state the theory of Affine Term Structure Model (ATSM) as the following:

Theorem 2 (Theory of Affine Term Structure Model)**.** *Assume* **(A2)** *in an arbitrage–free system. Then there are constants* $r_k, p_k^i, \sigma_k^{ij} = \sigma_k^{ji}$ *for* $k = 0, \cdots, n, i, j = 1, \cdots, n$, *such that the functions* $A_0(\cdot), \cdots, A_n(\cdot)$ *are solutions of the* **Riccati** equations

$$\begin{cases} \dfrac{dA_k(s)}{ds} & = r_k - \displaystyle\sum_{i=1}^n p_k^i A_i(s) - \frac{1}{2} \sum_{i,j=1}^n \sigma_k^{ij} A_i(s) A_j(s) \qquad \forall\, s \in [0, T^{\max}), \\ A_k(0) & = 0, \qquad\qquad k = 0, \cdots, n. \end{cases} \tag{2.2.4}$$

If $A_k(\cdot), A_i(\cdot)A_j(\cdot), k = 0, \cdots, n, i = 1, \cdots, n, j = 1, \cdots, i,$ *are linearly independent, then*

$$\frac{dZ_t^T}{Z_t^T} = R_t\, dt - \sum_{i=1}^n A_i(T-t) \left\{ P_t^i dt + dX_t^i \right\},$$

$$R_t = \sum_{k=0}^n r_k X_t^k, \qquad P_t^i = \sum_{k=0}^n p_k^i X_t^k, \qquad \frac{\mathrm{Cov}(dX_t^i, dX_t^j)}{dt} = \sum_{k=0}^n \sigma_k^{ij} X_t^k \quad \left(X_t^0 \equiv 1\right).$$

The proof is given in the Appendix A.2.

Remark 2.2.2. (1) Here (2.2.4) holds regardless of the independency of $A_1(\cdot), \cdots, A_n(\cdot)$.

Indeed, if $A_1, \cdots, A_n$ are linearly dependent, say $A_n = \sum_{i=1}^{n-1} c^i A_i$ for some constants $c^1, \cdots, c^{n-1}$, then setting $\hat{X}_t^i = X_t^i + c^i X_t^n$ we have $\sum_{i=1}^n A_i(s) X_t^i = \sum_{i=1}^{n-1} A_i(s) \hat{X}_t^i$; namely, (2.2.3) is an $(n-1)$-factor model.

(2) In general (2.2.4) does not have a global, i.e. for all $s \in [0, \infty)$, solution. We introduce $T^{\max} \in (0, \infty]$ to denote the longest terms of bond of interest.

(3) In ATSM (2.2.3), we call $X_t^1, \cdots, X_t^n$ **factors** and $A_0(\cdot), \cdots, A_n(\cdot)$ **loads**. All loads are uniquely determined by $\{r_k, \{p_k^i, \{\sigma_k^{ij}\}_{j=1}^i\}_{i=1}^n\}_{k=0}^n$, so there are a total of $[1 + \frac{n}{2}](n+1)^2$ parameters. In particular, when $n = 2$, there are a total of 18 parameters.

2.2.3 Linear Term Structure Model

In (2.2.3), $A_0(T-t)$ is the expectation of $-\log Z_t^T$ when the mean of each factor X_t^i, $i=1,\cdots,n$, is assumed to be zero. Empirical study shows that mean returns of securities are hard to measure and moving averages experience large oscillations. Partially for this reason and partially for simplicity, here we propose a linear term structure model, **LTSM** for short. We assume the following:

(A3) *The price Z_t^T of the T-bond at time t satisfies*

$$\log \frac{1}{Z_t^T} = \sum_{i=1}^n L_i(T-t)\, F_t^i \quad \forall\, t \geqslant 0, T \in [t, t+T^{\max}), \tag{2.2.5}$$

where $L_1(\cdot),\cdots,L_n(\cdot)$ are differentiable functions defined on $[0,T^{\max})$ and $\{(F_t^1,\cdots,F^n)\}$ is an Itô process with a positive definite matrix $(\mathbb{E}[F_\tau^i F_\tau^j])_{n\times n}$ for some $\tau \geqslant 0$.

Theorem 3 (Theory of the Linear Term Structure Model)**.** *Assume* **(A3)** *in an arbitrage–free environment. Then there are constants $r_k, p_k^i, \sigma_k^{ij} = \sigma_k^{ji}$ for $k,i,j = 1,\cdots,n$ such that the functions $L_1(\cdot),\cdots,L_n(\cdot)$ are solutions of the* **Riccati** equations

$$\begin{cases} \dfrac{dL_k(s)}{ds} & = r_k - \displaystyle\sum_{i=1}^n p_k^i L_i(s) - \frac{1}{2}\sum_{i,j=1}^n \sigma_k^{ij} L_i(s)L_j(s) \qquad \forall\, s \in [0,T^{\max}), \\ L_k(0) & = 0, \qquad\qquad k=1,\cdots,n. \end{cases} \tag{2.2.6}$$

If $L_i(\cdot), L_i(\cdot)L_j(\cdot), i=1,\cdots,n, j=1,\cdots,i$, are linearly independent, then

$$\frac{dZ_t^T}{Z_t^T} = R_t\, dt - \sum_{i=1}^n L_i(T-t)\left\{P_t^i dt + dF_t^i\right\}, \tag{2.2.7}$$

$$R_t = \sum_{k=1}^n r_k F_t^k, \qquad P_t^i = \sum_{k=1}^n p_k^i F_t^k, \qquad \frac{\mathrm{Cov}(dF_t^i, dF_t^j)}{dt} = \sum_{k=1}^n \sigma_k^{ij} F_t^k\,.$$

The proof is similar to that for Theorem 2 ATSM in the Appendix A.2 and therefore is omitted.

Remark 2.2.3. (1) In LTSM, the loads, $L_1(\cdot),\cdots,L_n(\cdot)$, are uniquely determined by the parameters $\{r_k, \{p_k^i, \{\sigma_k^{ij}\}_{j=1}^i\}_{i=1}^n\}_{k=1}^n$ so there are a total of $n(n+1)(1+n/2)$ parameters. When $n=2$, there are a total of 12 parameters.

(2) When $F_t^1 \equiv 1$, LTSM becomes an $(n-1)$-factor ATSM model.

(3) Introducing

$$m_t^i = \mathbb{E}[F_t^i], \qquad X_t^i = F_t^i - m_t^i, \quad L_0(s,t) := \sum_{i=1}^n L_i(s) m_t^i, \quad A_i(s) = L_i(s),$$

we can write (2.2.5) as

$$\log \frac{1}{Z_t^T} = L_0(T-t,t) + \sum_{k=1}^m A_i(T-t)\, X_t^i\ .$$

Thus, **LTSM** *generalizes* **ATSM** *by allowing time dependent mean returns.* When $\{(F_t^1, \cdots, F_t^m)\}_{t\in\mathbb{R}}$ is stationary, m_t^k does not depend on t, so $L_0(s,t)$ depends only on s and LTSM is a special ATSM in which $A_0(\cdot)$ is a linear combination of $A_1(\cdot), \cdots, A_n(\cdot)$.

2.2.4 The Black–Scholes Pricing

A Black Scholes theory evaluates prices of security derivatives based on no arbitrage and Itô calculus. We focus on a security derivative being the arrangement at time t of a payment P_T at a future time $T > t$. The payment P_T can be calculated after the observation at time T of values Z_T^{T+s} for all $s \in (0, T^{\max})$; that is, P_T is a functional of $(Z_T^{T+s})_{s\in(0,T^{\max})}$.

Now we consider LTSM. Assume that the model parameters $r_k, p_k^i, \sigma_k^{ij}$ for $k,i,j = 1, \cdots, n$ are all known and that the model is **irreducible** in the sense that the solutions $L_1(\cdots), \cdots, L_n(\cdot)$ of the Riccati equations (2.2.6) are linearly independent. Then one can find positive constants $s^1, \cdots, s^n$ such that the matrix

$$\mathbf{L}(s^1, \cdots, s^n) := \begin{pmatrix} L_1(s^1) & \cdots & L_1(s^n) \\ \vdots & \ddots & \vdots \\ L_n(s^1) & \cdots & L_n(s^n) \end{pmatrix} \tag{2.2.8}$$

is invertible. Consequently (2.2.5) implies that the values $F_t^1, \cdots, F_t^n$ of all factors at time t can be calculated from observations $Z_t^{t+s_1}, \cdots, Z_t^{t+s_n}$ of the bond prices via formula

$$(F_t^1,\ \cdots,\ F_t^n) = -\Big(\log Z_t^{t+s^1},\ \cdots\ , \log Z_t^{t+s^n}\Big)\mathbf{L}^{-1}(s^1, \cdots, s^n). \tag{2.2.9}$$

These values of factors, in turn, provide the time t prices of all bonds via (2.2.5). Thus,

all factors in an irreducible LTSM model are adapted to a natural filtration.

That a future payment P_T depends only on $(Z_T^{T+s})_{s\in(0,T^{\max})}$ is equivalent to say that P_T depends only on $(F_T^1, \cdots, F_T^n)$. Hence, we can assume that there exists a function Φ of $z = (z^1, \cdots, z^n) \in \mathbb{R}^n$ such that

$$P_T = \Phi(z^1, \cdots, z^n)\Big|_{(z^1,\cdots,z^n)=(F_T^1,\cdots,F_T^n)}. \tag{2.2.10}$$

We assume that L_i, L_iL_j for $i = 1, \cdots, n, j = 1, \cdots, i$ are linearly independent so the second assertion of Theorem 3 holds. We introduce the following functions:

$$\Big(\sigma^{ij}(z), P^i(z), R(z)\Big) = \sum_{k=1}^{n} \Big(\sigma_k^{ij},\ p_k^i, r_k\Big) z^k \qquad \forall\, z = (z^1, \cdots, z^n) \in \mathbb{R}^n.$$

Theorem 4 (Black– Scholes Pricing)**.** *Assume the* **LTSM** *in an arbitrage-free system and consider a security derivative with payoff P_T at time T where P_T is given by (2.2.10).*

Let $\Omega \subset \mathbb{R}^n$ be the smallest set such that the probability that $(F_t^1, \cdots, F_t^n) \in \Omega$ is one for every $t \leqslant T$. Assume that there exists a function $V : \Omega \times (T - T^{\max}, T] \to \mathbb{R}$ such that V is regular enough for the Itô formula to hold for $dV(F_t^1, \ldots, F_t^n, t)$ and

$$\begin{cases} \dfrac{\partial V}{\partial t} + \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\sigma^{ij}}{2} \frac{\partial^2 V}{\partial z^i \partial z^j} = RV + \sum_{k=1}^{n} P^k \frac{\partial V}{\partial z^k} & \text{in } \ \Omega \times (T - T^{\max}, T], \\ V(\cdot, T) = \Phi(\cdot) & \text{on } \ \Omega \times \{T\}. \end{cases} \tag{2.2.11}$$

Then at any time $t \in (T-T^{\max}, T]$, the value of the security derivative is $V(F_t^1, \cdots, F_t^n, t)$.

The proof will be given in the appendix A.3.

Remark 2.2.4. (1) For an n-factor ATSM, the same Theorem holds with

$$\Big(\sigma^{ij}(z), P^i(z), R(z)\Big) = \sum_{k=0}^{n} \Big(\sigma_k^{ij},\ p_k^i, r_k\Big) z^k\Big|_{z^0=1} \qquad \forall\, z = (z^1, \cdots, z^n) \in \mathbb{R}^n.$$

Also we can let $\Omega = \{1\} \times \hat{\Omega}$ to treat an (n-1)-factor ATSM as an LTSM with $F_t^1 \equiv 1$.

(2) One can check that when $\Phi \equiv 1$, (2.2.11) admits a solution

$$V(z,t) = \exp\Big(-\sum_{k=1}^{n} L_k(T-t) z^k\Big) \quad \forall t \leqslant (T - T^{\max}, T], z \in \mathbb{R}^n.$$

This gives the price $V(F_t, t) = e^{-\sum_{k=1}^{n} L_k(T-t)F_t^k}$ of the underlier T-bond.

(3) Since (2.2.11) is linear, it admits at most one regular (i.e. Itô formula applies to $dV(F_t,t)$) solution. Indeed, suppose there are two regular solutions, say V_1 and V_2. Then $W := V_2 - V_1$ is a regular solution with $\Phi \equiv 0$. Consequently, $e^{-\sum_{k=1}^{n} L_k(T-t)F_t^k} + W(F_t, t)$ is the price of T-bond. Hence, we must have $W(\cdot, t) \equiv 0$ in the state space Ω.

(4) The partial differential equation in (2.2.11) depends only on the functions $R(z), (\sigma^{ij}(z))_{n\times n}, (P^i(z))_{n\times 1}$, which in turn are completely determined by parameters $\{r_k, \{p_k^i\}_{i=1}^n, \{\{\sigma_k^{ij}\}_{i=1}^n\}_{j=1}^i\}_{k=1}^n$. Thus, Theorem 4 states in a sense that in pricing security derivatives, it is sufficient to find all these coefficients. There is little need to find precise models for the stochastic process $\{(F_t^1, \cdots, F_t^n)\}_{t\geqslant 0}$. This is the celebrated well-known advantage of the Black-Scholes theory since in general it is very hard to select a particular process that fits empirically the behavior of factors. On the other hand, as we shall show in subsequent sections, we can determine empirical values of these constants, thereby completely pinning down the partial differential equation for the price function.

(5) To establish the well-posedness (existence, uniqueness, and continuous dependence on parameters) of problem (2.2.11), additional qualitative (not quantitative) information on the process $\{(F_t^1, \cdots, F_t^n)\}$ is needed. We shall elaborate this topic

in the next section.

2.2.5 Econometrics

The return of a T-bond is in general quoted by

$$y_t^T := \frac{1}{T-t} \log \frac{1}{Z_t^T} \qquad \Big(\iff Z_t^T = \exp(-[T-t]y_t^T) \Big).$$

We call $s = T - t$ the **time-to-matutity**, **duration** or **term** of the bond and y_t^{t+s} the **yield-to-maturity** or simply the yield of bond with duration s. For fixed t, the curve $\{(s, y_t^{t+s}) \mid s \geqslant 0\}$ is called a **yield-to-maturity curve**. A prescription of a time t yield-to-maturity curve is equivalent to a prescription of time t prices of all bonds.

Under the ATSM, the yield is given by

$$y_t^{t+s} = a_0(s) + \sum_{i=1}^{n} a_k(s) X_t^k, \qquad a_i(s) = \frac{A_i(s)}{s},$$

and under LTSM,

$$y_t^{t+s} = \sum_{k=1}^{n} \ell_k(s) F_t^k = \ell_0(s) + \sum_{k=1}^{n} \ell_k(s) \Big(F_t^k - m^k \Big), \ell_0(s) := \sum_{k=1}^{n} \ell_k(s) m^k, \ell_k(s) = \frac{L_k(s)}{s}.$$

From econometric point of view, here we shall use a large collection of empirical data $\{y_{t_i}^{t_i+s^i} \mid i = 1, \cdots, N, j = 1, \cdots, m\}$ to find empirical loads $a_i(\cdot), \ell_k(\cdot)$ and factors X_t^i, F_t^k. Also we find optimal parameters appeared in the Riccati equations so that the solutions match the empirical ones. It is important to notice that knowing all parameters of LTSM (or ATSM) allows us to calculate the prices of security derivatives by solving the Black-Scholes equation (2.2.11).

2.3 CLARIFICATIONS ABOUT THE BLACK-SCHOLES EQUATIONS

From the theory of partial differential equation point of view, problem (2.2.11) is not complete since (i) the prescription of Ω is vague and (ii) conditions of V on the boundary $\partial\Omega \times (T - T^{\max}, T)$ is not prescribed. We shall now address these two questions. We focus our discussion on LTSM, as that for ATSM is analogous.

2.3.1 The State Space

We call $F_t := (F_t^1, \cdots, F_t^n)^\top$ ($^\top$ stands for transpose) state variables. The state space Ω is the set in $\mathbb{R}^n$ that can be reached by the state variables:

$$\Omega := \underset{t \geqslant 0}{\cup} \Omega_t, \qquad \Omega_t := \bigcap \Big\{ A \mid \text{Probability}(F_t \in A) = 1 \Big\}.$$

For simplicity we assume that $\Omega = \Omega_t$ for all $t > 0$.

Recall that $\{F_t\}$ is assumed to be an Itô process. It means that there exist an $n \times 1$ vector function $\mathbf{b}(z, t)$ and an $n \times n$ matrix function $\mathbf{a}(z, t)$ in certain class such that

$$dF_t = \mathbf{b}(F_t, t)\, dt + \mathbf{a}(F_t, t)\, dW_t, \qquad F_t \in \Omega \tag{2.3.1}$$

where $\{W_t\} = \{(W_t^1, \cdots, W_t^n)^\top\}$ is the standard Wiener process. This implies that

$$\text{Cov}(dF_t, dF_t^\top) = \mathbf{a}(F_t, t)\, \text{Cov}(dW_t, dW_t^\top)\mathbf{a}^\top(F_t, t) = \mathbf{a}(F_t, t)\mathbf{a}^\top(F_t, t)\, dt,$$
$$\mathbf{a}(z, t)\mathbf{a}^\top(z, t) = \sigma(z) := (\sigma^{ij}(z))_{n\times n} := \Big(\textstyle\sum_{k=1}^n \sigma_k^{ij} z^k \Big)_{n\times n}.$$

It then follows that

$$\Omega \subset D_1 := \{ z \in \mathbb{R}^n \mid \sigma(z) \geqslant 0 \} \qquad \forall\, z \in \Omega.$$

Here $\sigma(z) \geqslant 0$ means that the matrix $\sigma(z)$ is semi-positive definite. Also, since the bond price $Z_t^T = \exp(-\sum_{k=1}^n L_i(T-s)F_t^k)$ cannot exceed 1, we need

$$\Omega \subset D_2 = \Big\{(z^1, \cdots, z^n) \mid \sum_{k=1}^n L_k(s) z^k \geqslant 0 \quad \forall\, s \in [0, T^{\max})\Big\}.$$

Hence, *the maximum state space we could take is* $\Omega := D_1 \cap D_2$.

In [42], the state space is taken to be

$$D := \{z \mid \sigma^{11}(z) \geqslant 0, \cdots, \sigma^{nn}(z) \geqslant 0\}.$$

Under a non-degeneracy condition, Duffie and Kan [42] demonstrated that there exist constant non-singular matrix $\Sigma_{n\times n}$ and row vectors $\beta_1, \cdots, \beta_n$ such that

$$\sigma(z) = \Sigma \;\; \mathrm{diag}(\beta_1 z, \cdots, \beta_n z) \;\; \Sigma^\top. \tag{2.3.2}$$

From empirical point of view, this imposes quite a number of restrictions on the parameters.

2.3.2 The Risk-Neutral Measure.

In the benchmark work [42], Duffie and Kan carried out all important ingredients of the Affine Term Structure Model, except that the vector function $\mathbf{b}$ in (2.3.1) is not fully attended since only the risk-neutral measure is important in evaluating security derivatives. There is basically no restriction on $\mathbf{b}$ so we hope we can take specific $\mathbf{b}$ so that the state space Ω can take our favorable choice.

In the classical ATSM model [42], the state variables $X_t = (X_t^1, \cdots, X_t^n)^\top$ satisfy

$$dX_t = P(X_t)\, dt + \mathbf{a}(X_t) dW_t^*, \qquad X_t \in D \tag{2.3.3}$$

where $\{W_t^*\}$ is a standard n-dimensional Winner process whose martingale-measure is the risk-neutral measure. Since both $P(z)$ and $\mathbf{a}(z)\mathbf{a}^\perp(z)$ are affine function of z, working on $\Omega = D$ leads to a number of simplifications as well as restrictions on the parameters $(r_k, p_k^i, \sigma_k^{ij})$. It should be noted that under (2.3.3), $\{X_t\}$ is adapted to the risk-neutral measure of $\{W_t^*\}$.

In empirical estimation of parameters, it may not be compatible to start with (2.3.3) and (2.2.3), since on the one hand, one measures $\{X_t\}$ through a physical probability measure but on the other hand, one uses an equation under risk-neutral measure.

2.3.3 The physical Probability Measure

In empirical estimation of parameters, one can start with (2.3.1) and (2.2.5), since all stochastic processes are observed under physical probability, not risk neutral probability. Also, in pricing a security derivative by the formula $V(F_t, t)$, the value F_t is empirical, i.e. under (2.3.1).

Note from our proofs of Theorems 1– 3 that there is no structural restriction on $\mathbf{b}(z, t)$ in (2.3.1) for the whole theory of LTSM to be consistent. This gives us extra room in choosing the state space Ω, in contrast to that of (2.3.3), where $P(z)$ has to be linear in z.

2.3.4 Probability Density

The equation $dF_t = \mathbf{b}(F_t, t)dt + \mathbf{a}(F_t, t)dW_t$ has its physical definition domain $\mathbb{R}^n$ and artificial definition domains such as that in (2.3.1). The range of the solution under the physical definition domain may happen to be the artificial one; nevertheless, using artificial domain introduces flexibilities. Working on the natural filtration of F_t, we can ignore the filtration of $\{W_t\}$, so directly deal with probability density may be sufficient.

Consider (2.3.1). For $s \geqslant t$, denote by $\rho(x, t; z, s)$ the probability density of F_s

under condition $F_t = x$. Then, as a function of (z, s), $\rho = \rho(x, t; z, s)$ satisfies

$$\begin{cases} \dfrac{\partial \rho}{\partial s} = \displaystyle\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2(\sigma^{ij}\rho)}{2\,\partial z^i z^j} - \sum_{i=1}^{n} \frac{\partial(b^i \rho)}{\partial z^i} & \text{in } \ \Omega \times (t, \infty), \\ \\ \lim_{s \searrow t} \rho(x, t; \cdot, s) = \delta(\cdot - x) \end{cases} \tag{2.3.4}$$

where $(b^i)_{n\times 1} = \mathbf{b}(z, s)$, $(\sigma^{ij})_{n\times n} = \mathbf{a}(z, s)\mathbf{a}^\top(z, s)$ and $\delta(\cdot - x)$ is the Dirac mass at x.

Equation (2.3.4) may not be complete since boundary conditions on $\partial\Omega$ may be needed. If we regard F_t as the position of a particle, we have to specify what is the subsequent motion if the particle hits the boundary $\partial\Omega$. For illustration, we consider the following cases:

(i) The probability that F_t hits the boundary is zero. Then (2.3.4) is complete.

Example 1. Let $\Omega = (0, \infty)$ and $S_t = e^{W_t}$. Then $dS_t = S_t[\frac{1}{2}dt + dW_t]$. The density of S_s is $\rho(z, s) = e^{-(\ln z)^2/(2s)}/[z\sqrt{2\pi s}]$. We have two automatically fulfilled boundary conditions

$$\lim_{z \searrow 0} \frac{\partial \rho(z, s)}{\partial z} = 0, \qquad \lim_{z \searrow 0} \rho(z, s) = 0 \quad \forall\, s > 0.$$

(ii) Suppose that F_t does hit the boundary $\partial\Omega$ and that $\partial\Omega$ is a hard wall, so particles bounce back. This corresponds to supply (2.3.4) with the boundary condition

$$\lim_{y \leftarrow z \in \Omega} \sum_{i=1}^{n} n^i(y) \Big(\sum_{j=1}^{n} \frac{\partial(\sigma^{ij}(z)\rho(x, t; z, s))}{2\,\partial z^i} - b^i(z, s)\rho(x, t; z; s) \Big) = 0 \,\forall\, y \in \partial\Omega, s > t \tag{2.3.5}$$

where $\mathbf{n}(y) = (n^1(y), \cdots, n^n(y))$ is the unit exterior normal to $\partial\Omega$ at $y \in \partial\Omega$.

Example 2. Let $\Omega = (0, \infty)$ and $x_t = (W_t)^2$. Then $dx_t = dt + 2\sqrt{x_t}\,\mathrm{sgn}(W_t)\, dW_t$, $x_t \geqslant 0$. We know x_s obeys a χ square distribution with density $\rho(z, s) = e^{-z/(2s)}/(\sqrt{2\pi z s})$. For $\sigma(z) = [2\sqrt{z}]^2 = 4z$ and $b(z, s) \equiv 1$, we have

$$\lim_{z \searrow 0} \Big\{ \frac{\partial[\sigma(z)\rho(z, s)]}{2\,\partial z} - b(z, s)\rho(z, s) \Big\} = 0 \quad \forall\, s > 0.$$

(iii) Suppose F_t hits the wall $\partial\Omega$ and that the wall is soft so once the particle hits the wall, it stays there. Then the probability density of F_s under condition $F_t = x$ is $\hat{\rho}(x; \cdot, s) = \rho(x, t; \cdot, s) + \rho^*(x; \cdot, s)\delta_{\partial\Omega}$ where $\delta_{\partial\Omega}$ is the Dirac measure concentrated on

$\partial\Omega$, ρ is the solution of (2.3.4) with boundary condition $\rho = 0$ on $\partial\Omega \times (0, \infty)$ and ρ^* is the percentages of particles landed on the wall and is given by

$$\rho^*(x, t; z, s) = \int_t^s \sum_{i,j=1}^n n^i(z) \frac{\partial(a^{ij}(z)\rho(x; z, \tau))}{2\,\partial z^i} d\tau \qquad \forall\, z \in \partial\Omega, s > t.$$

Example 3. Set $\Omega = [0, \infty)$ and $x_t = [1 + W_t]^2$ if $\min_{s\in[0,t]} W_s > -1$ and $X_t = 0$ otherwise. Then $dx_t = dt + 2\sqrt{x_t}\, dW_t$, $x_t \geqslant 0$. The density of x_s is $\hat{\rho}(\cdot, s) = \rho(\cdot, s) + \rho^*(s)\delta(\cdot)$ where

$$\rho(z, s) = \frac{e^{-(1-\sqrt{z})^2/(2s)}}{2z\sqrt{2\pi s}}\Big(1 - e^{-2\sqrt{z}/s}\Big), \qquad \rho^*(s) = \int_0^s \frac{1}{\sqrt{2\pi\tau}\ \tau} e^{-1/(2\tau)} d\tau.$$

2.3.5 Solutions of the Black Scholes Equation.

We use Green's function to represent the solution. Let $\rho = \rho(x; z, s)$ be solution, for $(z, s) \in \Omega \times [0, T^{\max})$, of

$$\begin{cases} \dfrac{\partial\rho}{\partial s} = \dfrac{1}{2}\displaystyle\sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2[\sigma^{ij}\rho]}{\partial z^i \partial z^j} + \sum_{i=1}^n \frac{\partial[P^i\rho]}{\partial z^i} - R\rho & \text{in } \ \Omega \times (0, T^{\max}), \\ \rho(x; \cdot, 0) = \delta(x - \cdot) & \text{on } \ \Omega \times \{0\}, \\ \sum_{i=1}^n n^i \Big\{ \sum_{j=1}^n \frac{\partial[\sigma^{ij}\rho]}{2\,\partial z^j} + P^i\rho \Big\} = 0 & \text{on } \ \partial\Omega \times (0, T^{\max}). \end{cases} \tag{2.3.6}$$

Below subscripts are partial derivatives; a repeated index implies an omitted summation over the index from 1 to n. Since $\rho(x; z, 0)$ is a delta function, we have

$$\begin{aligned} &\int_\Omega \Phi(z)\rho(x; z, T - t)dz - V(x, t) = \int_\Omega V(z, t + s)\rho(x; z, s)\Big|_{s=0}^{s=T-t} dz \\ = &\int_\Omega \int_0^{T-t} \frac{\partial}{\partial s}\Big[V(z, t + s)\rho(x; z, s)\Big] dz ds = \int_\Omega \int_0^{T-t} [V_s\rho + V\rho_s] dz ds \\ = &\int_\Omega \int_0^{T-t} \Big\{ \Big[RV + P^i V_{z^i} - \tfrac{1}{2}\sigma^{ij} V_{z^i z^j}\Big]\rho + V\Big[\tfrac{1}{2}(\sigma^{ij}\rho)_{z^i z^j} + (P^i\rho)_{z^i} - R\rho\Big]\Big\} ds dz \\ = &\int_0^{T-t} \int_\Omega \Big\{ \tfrac{1}{2}[V(\sigma^{ij}\rho)_{z^j}]_{z^i} - \tfrac{1}{2}[V_{z^i}\sigma^{ij}\rho]_{z^j} + [VP^i\rho]_{z^i} \Big\} dz ds \\ = &\int_0^{T-t} \int_{\partial\Omega} n^i \Big\{ \tfrac{1}{2}V(\sigma^{ij}\rho)_{z^j} - \tfrac{1}{2}V_{z^j}\sigma^{ji}\rho + P^i V\rho \Big\} dS ds = -\int_0^{T-t} \int_{\partial\Omega} \frac{\rho}{2}\ V_{z^j}\sigma^{ji} n^i dS ds \end{aligned}$$

where dS is the surface element of $\partial\Omega$. Thus, we have the **Green's identity:**

$$V(x,t) = \int_\Omega \Phi(z)\rho(x;z,T-t)\,dz + \frac{1}{2}\int_t^T\int_{\partial\Omega} \rho(x;z,T-s)\nabla^\top V(z,s)\sigma(z)\mathbf{n}(z)dSds \tag{2.3.7}$$

where $\nabla^\top = (\partial/\partial z^1, \cdots, \partial/\partial z^n)$. The above calculation can be made rigorous so we have the following:

Theorem 5. *Let Ω be an open domain in which $\sigma(z) > 0$. Assume that $\sigma\mathbf{n} = \mathbf{0}$ on $\partial\Omega$ and (2.3.6) admits a (weak) solution. Then for each bounded Φ, the Black-Scholes system (2.2.11) admits a solution and the solution is unique in the class of functions with bounded derivatives. The unique solution is given by (2.3.7) with the boundary integral removed.*

In conclusion, if F_t does hit the boundary of the state space with positive probability, we need not abandon the affine term structure model; instead, we ask for more information about the behavior of the state variables after they hit the boundary.

Remark 2.3.1. (1) It is expected that under the affine structure of the function (σ^{ij}, P^i, R) and under the assumption that $\sigma > 0$ in Ω and $\sigma\,\mathbf{n} = \mathbf{0}$ on $\partial\Omega$, (2.2.11) admits a unique (weak) solution. We shall get into details here.

(2) When $\Omega = (0,\infty)$, $\sigma\,\mathbf{n} = 0$ on $\partial\Omega$ is equivalent to $\sigma(0) = 0$.

(3) The boundary condition in (2.3.6) should be interpreted as (2.3.5), since ρ may not be even bounded; See **Example 2** above. If, on the other hand, ρ is differentiable up to the boundary, then $\sigma\,\mathbf{n} = \mathbf{0}$ on $\partial\Omega$ implies that the boundary condition in (2.3.6) is equivalent to $\rho = 0$ on $\partial\Omega \times (0, T^{\max})$.

(3) The condition $\sigma\,\mathbf{n} = \mathbf{0}$ on $\partial\Omega$ is fulfilled so the Black-Scholes equation admits a unique solution, under the Condition A part (b) of Duffie and Kan [42, p387], which, referring to (2.3.2) and denoting $\beta_i\Sigma = (c_i^1, \cdots, c_i^n)$, states as follows: *If $c_i^j \neq 0$, then $\beta_j = \beta_i$.*

The proof goes as follows. In [42], $\Omega = D = \{z \mid \beta_i z > 0, i = 1, \cdots, n\}$. Since D is non-empty, $|\beta_i| > 0$ for all $i = 1, \cdots, n$. Suppose $z \in \partial\Omega$. Then $\beta_i z = 0$ for some

i. Since a normal of the hyperplane $\beta_i z = 0$ is $\beta_i^\top$, $\mathbf{n}(z)$ is parallel to $\beta_i^\top$. Hence,

$$\begin{aligned} & \sigma(z)\mathbf{n}(z) = \mathbf{0} \Leftrightarrow \sigma(z)\beta_i^\top = \mathbf{0} \Leftrightarrow \beta_i\sigma(z) = \mathbf{0} \Leftrightarrow \beta_i\Sigma \operatorname{diag}(\cdots)\,\Sigma^\top = \mathbf{0} \\ \Leftrightarrow \quad & \beta_i\Sigma \operatorname{diag}(\beta_1 z, \cdots, \beta_n z) = \mathbf{0} \Leftrightarrow (c_i^1\beta_1 z, \cdots, c_i^n\beta_n z) = \mathbf{0}. \end{aligned}$$

Now if $c_i^j \neq 0$, then $\beta_j = \beta_i$ so that $c_i^j\beta_j z = c_i^j\beta_i z = 0$. Hence, $\sigma(z)\mathbf{n}(z) = \mathbf{0}$ for every $z \in \partial\Omega$.

2.4 FACTORS AND LOADS

One of the central issue in term structure model is the characterization of the stochastic process $\{X_t^i\}$. We shall define them in terms of statistical common factors. Here, by statistical, it means factors are obtained from the given random variables themselves. This section explains the basic theory that we use.

2.4.1 Principal Component Analysis

Let $(\mathbf{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space. We use $\dim(V)$ to denote the dimension of a subspace V of $\mathbf{H}$. For $f, f^1, \cdots, f^n$ in $\mathbf{H}$ we denote

$$\operatorname{dist}(f, V) = \inf_{h\in V} \|f - h\|, \quad (\{f_1, \cdots, f_n\}) := \{c_1 f^1 + \cdots + c_n f^n \mid c_1, \cdots, c_n \in \mathbb{R}\}.$$

In many applications, one runs into a large collection of random variables and would like to model them by a space of small dimension. This leads to the following definition.

Definition 1. *Let $\xi^1, \cdots, \xi^N$ be points in $\mathbf{H}$. Set $\xi = \{\xi^1, \cdots, \xi^N\}$.*

(1) *A* **principal subspace** *of ξ is a subspace V of $\mathbf{H}$ satisfying*

$$\sum_{i=1}^N \operatorname{dist}^2(\xi_i, V) = \min_{\dim(W)=\dim(V)} \sum_{i=1}^N \operatorname{dist}^2(\xi_i, W).$$

(2) *A set of* **principal components** *of* ξ *is an ordered orthonormal set* $\{F^1, \cdots, F^n\}$ *in* $\mathbf{H}$ *such that for each* $k = 1, \cdots, n$, $\{F^1, \cdots, F^k\}$ *is a principal subspace of* ξ.

It is easy to see the following:

(i) If $\{F_1, \cdots, F^n\}$ is a set of principal components, so is $\{F^1, \cdots, F^k\}$ for each $k = 1, \cdots, n$.

(ii) If V is a principal subspace of ξ and $\dim(V) \leqslant \dim((\xi))$, then $V \subset (\xi)$.

Principal components can be found by the following; see, e.g. [89] and references therein.

Theorem 6. *Let* $\xi^1, \cdots, \xi^N$ *be points in a Hilbert space* $(\mathbf{H}, \langle\cdot,\cdot\rangle)$ *and* $\{\lambda_i\}_{i=1}^N$, *arranged in decreasing order, be a complete set of eigenvalues of* $\mathbf{C} := (\langle\xi^i, \xi^j\rangle)_{N\times N}$. *Let* $K = \dim(\{\xi^1, \cdots, \xi^N\})$. *Then* $\{F^1, \cdots, F^K\}$ *is a set of principal components of* $\{\xi^1, \cdots, \xi^N\}$ *if and only if there exist row vectors* $\mathbf{e}_1, \cdots, \mathbf{e}_K$ *in* $\mathbb{R}^N$, $\mathbf{e}_k = (e_k^1, \cdots, e_k^N)$, *such that*

$$\mathbf{e}_k\mathbf{C} = \lambda_k\, \mathbf{e}_k, \qquad \mathbf{e}_k \cdot \mathbf{e}_l = \delta_{kl}, \qquad F^k = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^N \xi^i\, e_k^i \quad \forall\, k, l = 1, \cdots, K.$$

In addition,

$$\min_{\dim(V)=n} \sum_{i=1}^N \mathrm{dist}^2(\xi^i, V) = \sum_{i=1}^N \Big\| \xi^i - \sum_{k=1}^n \langle\xi^i, F^k\rangle F^k \Big\|^2 = \sum_{k=n+1}^N \lambda_k \quad \forall\, n = 1, \cdots, K,$$

$$R_n := \frac{\sum_{i=1}^N \mathrm{dist}^2(\xi^i, \{F^1, \cdots, F^n\})}{\sum_{i=1}^N \|\xi^i\|^2} = \frac{\sum_{k=n+1}^N \lambda_k}{\sum_{k=1}^N \lambda_k}. \tag{2.4.1}$$

The proof involves elementary linear algebra and is given in the Appendix A.4.

Remark 2.4.1. (1) R_n is an indicator of the goodness of accommodating N points in an n-dimensional space.

(2) When each point has different importance, it is better to minimize $\sum \omega_i \mathrm{dist}^2(\xi^i, V)$

with certain weights $(\omega_1, \cdots, \omega_N)$. Then it is nice to observe the following identity:

$$\sum_{i=1}^{N} \omega_i \operatorname{dist}^2(\xi^i, V) = \sum_{i=1}^{N} \operatorname{dist}^2(\sqrt{\omega_i}\xi^i, V).$$

Thus the problem becomes the standard PCA on $\{\sqrt{w_1}\xi_1, \cdots, \sqrt{\omega_N}\xi_N\}$.

2.4.2 Principal Subspace of Yield Curves

Let $\mathbf{T} = \{t_i\}_{i=1}^N$ be historical trading dates and $\mathbf{S} = \{s^j\}_{j=1}^m$ be time-to-maturities of bonds. If necessary, we regard $\mathbf{T}$ as a column vector and $\mathbf{S}$ as a row vector. Let y_t^T be the time t yield of zero-coupon T-bond. We want to accommodate the N yield-to-maturity curves: $s \in \mathbf{S} \to y_t^{t+s}$, $t \in \mathbf{T}$, into a space of small dimension. For this, we introduce, for each $t \in \mathbf{T}$, a function

$$Y(t, \cdot) : \mathbf{S} \to \mathbb{R}, \qquad Y(t, s) = y_t^{t+s} \qquad \forall\, s \in \mathbf{S}.$$

Let $\omega : \mathbf{S} \to (0, \infty)$ be a positive function selected as weights. We use the inner product, for $\mathbf{H} := L^2(\mathbf{S})$,

$$\langle f, g \rangle = \sum_{s \in \mathbf{S}} \omega(s) f(s) g(s).$$

Thus, we would like to find an orthonormal set $\{\beta_1, \cdots, \beta_m\}$ in $L^2(\mathbf{S})$ such that

$$\sum_{t \in \mathbf{T}} \operatorname{dist}^2(Y(t, \cdot), \{\beta_1, \cdots, \beta_k\}) = \min_{\dim(V)=k} \sum_{t \in \mathbf{T}} \operatorname{dist}^2(Y(t, \cdot), V) \quad \forall\, k = 1, \cdots, m.$$

To solve the problem, we regard $Y(t, \cdot)$ as a row vector $Y(t, \mathbf{S}) = (Y(t, s^1), \cdots, Y(t, s^m))$, and $Y(\cdot, \cdot)$ as a matrix $\mathbf{Y} = Y(\mathbf{T}, \mathbf{S}) = (Y(t_i, s^j))_{N \times m}$. Set $W = \operatorname{diag}(\omega(s^1), \cdots, \omega(s^m))$. Then

$$\mathbf{C} := \Big(\langle Y(t_i, \cdot), Y(t_j, \cdot) \rangle \Big)_{N \times N} = \mathbf{Y} W \mathbf{Y}^\top .$$

Let $\{\lambda_k\}_{k=1}^N$ be all the eigenvalues of $\mathbf{C}$, arranged in decreasing order. Denote the orthonormal (row) eigenvectors associated with $\lambda_1, \cdots, \lambda_N$ by $\mathbf{e}_1, \cdots, \mathbf{e}_N$. Let $K = \operatorname{rank}(Y) \leqslant m$. According to Theorem 6, the principal components $\{\beta_1, \cdots, \beta_K\}$ are given by

$$\beta_i(\mathbf{S}) = \mathbf{e}_i \mathbf{Y} / \sqrt{\lambda_i} \qquad \forall\, i = 1, \cdots, K.$$

Since $\mathbf{C}$ is an $N \times N$ matrix and usually N is large, directly solving the eigenvalue problem for $\mathbf{C}$ is not efficient. We use the following fact:

$$\mathbf{e}_i \mathbf{C} = \lambda_i \mathbf{e}_i \quad \Rightarrow \mathbf{e}_i (\mathbf{Y} W \mathbf{Y}^\top)(\mathbf{Y} W^{\frac{1}{2}}) = \lambda_i \mathbf{e}_i (\mathbf{Y} W^{\frac{1}{2}}) \quad \Rightarrow \beta_i W^{\frac{1}{2}} [W^{\frac{1}{2}} \mathbf{Y}^\top \mathbf{Y} W^{\frac{1}{2}}] = \lambda_i \beta_i W^{\frac{1}{2}}.$$

Thus, $(\lambda_i, \beta_i W^{\frac{1}{2}})$ is the eigenpair of the $m \times m$ matrix $W^{\frac{1}{2}} \mathbf{Y} \mathbf{Y}^\top W^{\frac{1}{2}}$. Hence we obtain the following.

Table 1: Sample Daily Data of US Treasury Fixed–Term Bond Yields

Date	3mo	6mo	1yr	2yr	3yr	5yr	7yr	10yr	20yr	30yr
10/1/1993	2.98	3.11	3.35	3.84	4.18	4.72	5.03	5.34	6.12	5.98
			...	... ...	...	...				
10/3/1994	5.05	5.61	6.06	6.69	7.01	7.35	7.52	7.66	8.02	7.86
			...	... ...	...	...				
10/2/1995	5.53	5.64	5.65	5.82	5.89	5.98	6.10	6.15	6.61	6.48
			...	... ...	...	...				
10/1/1996	5.10	5.35	5.65	6.03	6.22	6.39	6.54	6.65	6.99	6.88
			...	... ...	...	...				
10/1/1997	5.10	5.27	5.44	5.75	5.83	5.93	6.05	6.04	6.38	6.33
			...	... ...	...	...				
10/1/1998	4.23	4.36	4.28	4.17	4.10	4.10	4.26	4.33	5.09	4.90
			...	... ...	...	...				
10/1/1999	5.16	5.32	5.47	5.83	5.93	6.00	6.23	6.06	6.55	6.19
10/2/2000	6.27	6.33	6.06	5.98	5.92	5.86	5.95	5.83	6.18	5.93
			...	... ...	...	...				
10/1/2001	2.37	2.37	2.47	2.82	3.18	3.90	4.33	4.55	5.39	5.38
			...	... ...	...	...				
10/1/2002	1.59	1.54	1.56	1.80	2.11	2.75	3.34	3.72	4.81	4.93
			...	... ...	...	...				
10/3/2003	0.95	1.00	1.13	1.47	1.93	2.84	3.40	3.96	4.92	5.00
			...	... ...	...	...				
10/3/2004	1.71	2.00	2.21	2.63	2.92	3.44	3.85	4.21	4.95	5.06
			...	... ...	...	...				
10/3/2005	3.61	4.02	4.09	4.21	4.23	4.25	4.31	4.39	4.67	4.58
			...	... ...	...	...				
12/29/2006	5.02	5.09	5.00	4.82	4.74	4.70	4.70	4.71	4.91	4.81
Mean	3.96	4.13	4.28	4.60	4.78	5.07	5.30	5.42	5.92	5.81
Variance	1.67	1.70	1.67	1.60	1.48	1.28	1.16	1.05	0.92	0.88
Skewness	−0.62	−0.63	−0.59	−0.50	−0.39	−0.12	0.03	0.28	0.37	0.49
Kurtosis	1.95	2.00	2.04	2.17	2.19	2.17	2.10	2.24	2.34	2.52

- **PCA.** Let $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_m)$ $(\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m)$ and E satisfy

$$EE^\top = E^\top E = \mathbf{I}_{m\times m}, \qquad W^{\frac{1}{2}}\mathbf{Y}^\top \mathbf{Y} W^{\frac{1}{2}} = E\Lambda E^\top.$$

Set $\beta_i(\mathbf{S})$ as the ith row of $\beta = E^\top W^{-\frac{1}{2}}$ and $f^j_{\mathbf{T}}$ as the jth column of $f = YW^{\frac{1}{2}}E\Lambda^{-\frac{1}{2}}$.

- **Principal Decomposition**. Note that $f\Lambda^{\frac{1}{2}}\beta = \mathbf{Y}$, so

$$y_t^{t+s} = Y(t,s) = \sum_{k=1}^m \sqrt{\lambda_k}\ f_t^k\ \beta_k(s) \quad \forall\, s \in \mathbf{S}, t \in \mathbf{T}.$$

- $\{\beta_1, \cdots, \beta_m\}$ **are principal components of the family** $\{Y(t, \bullet)\}_{t\in\mathbf{T}}$ **in** $L^2(\mathbf{S})$.

Indeed, $(\langle \beta_i, \beta_j\rangle)_{m\times m} = \beta W \beta^\top = \mathbf{I}_{m\times m}$, so $\{\beta_1, \cdots, \beta_m\}$ is orthonormal. Also $\Lambda^{-\frac{1}{2}} f^\top \mathbf{Y} = \beta$, so $\beta_i = (f^i)^\top \mathbf{Y}/\sqrt{\lambda_i}$. Finally, $f^\top(\mathbf{Y}W\mathbf{Y}^\top) = \Lambda f^\top$. Hence, by Theorem 6, $\{\ell_1, \cdots, \ell_m\}$ are principal components.

- $\{f^1, \cdots, f^m\}$ **are principal components of the family** $\{\sqrt{\omega(s)}\ Y(\bullet, s)\}_{s\in\mathbf{S}}$ **in** $L^2(\mathbf{T})$.

Here $L^2(\mathbf{T})$ is equipped with the inner product $\langle \phi, \psi\rangle := \sum_{t\in\mathbf{T}} \phi(t)\psi(t)$.

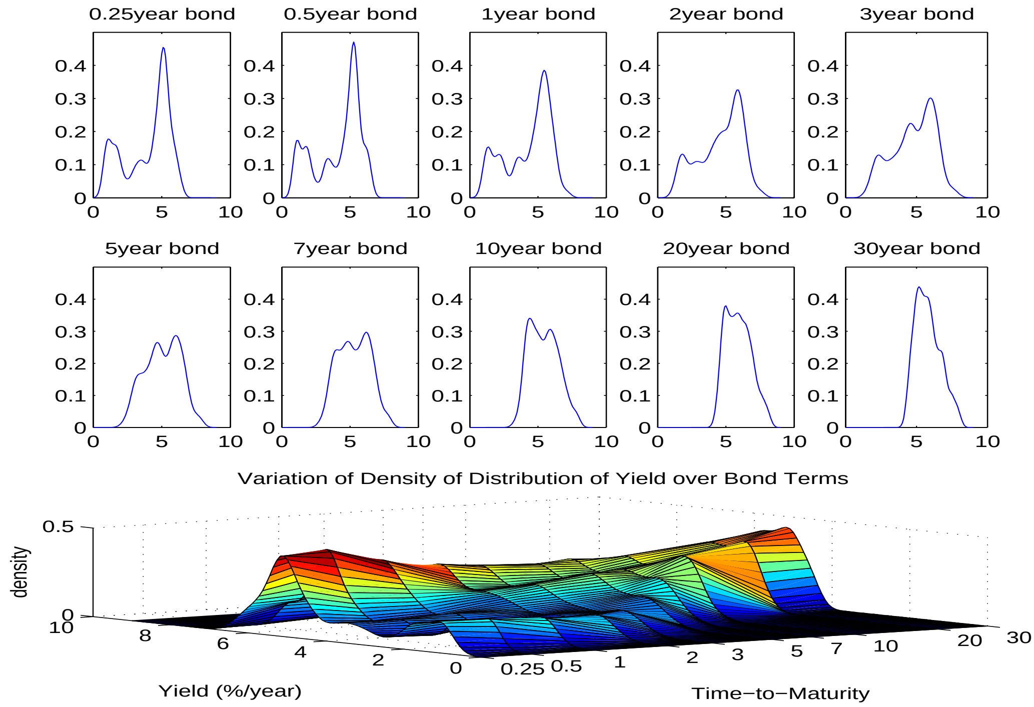
Indeed, $f^\top f = \mathbf{I}_{n\times n}$, so $\{f^1, \cdots, f^m\}$ is orthonormal. Also, $f = \mathbf{Y}W^{\frac{1}{2}}E\Lambda^{-\frac{1}{2}}$ so $f^i = (\mathbf{Y}W^{\frac{1}{2}})E^i/\sqrt{\lambda_i}$. Finally, $\hat{\mathbf{C}} := (\langle \sqrt{w(s^i)}Y(\bullet, s^i), \sqrt{w(s^j)}Y(\bullet, s^j)\rangle)_{m\times m} = W^{\frac{1}{2}}\mathbf{Y}^\top\mathbf{Y}W^{\frac{1}{2}}$ and $\hat{\mathbf{C}}E = E\Lambda$. Hence, by Theorem 6, $\{f^1, \cdots, f^m\}$ are principal components.

We call $\{\sqrt{N}f_t^1, \cdots, \sqrt{N}f_t^m\}$ the **factors** and call $\{\sqrt{\lambda_1/N}\ell_1(s), \cdots, \sqrt{\lambda_m/N}\ell_m(s)\}$ the **loads**, though, after scaling, both can be regarded as principal components, depending on the setting, i.e., regarding $\{y_t^{t+s}\}$ as an $\mathbf{S}$ family of functions of t or a $\mathbf{T}$ family of functions of s.

2.5 MODELLING THE US TREASURY BONDS

Based on the above theoretical framework, in this section we use historical data of US Government bonds of various maturities to find a complete affine term structure model (ATSM) and its general version, LTSM.

Figure 1: Empirical Densities of Distribution of Yields of Various Bonds

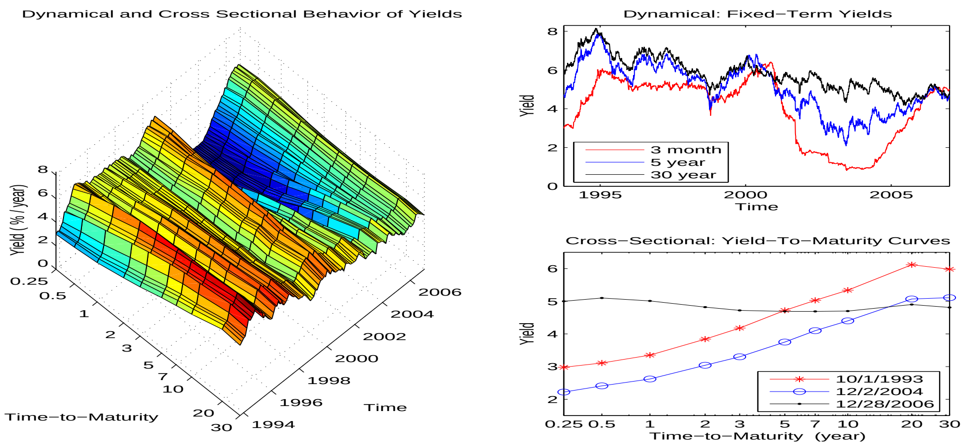


2.5.1 Data

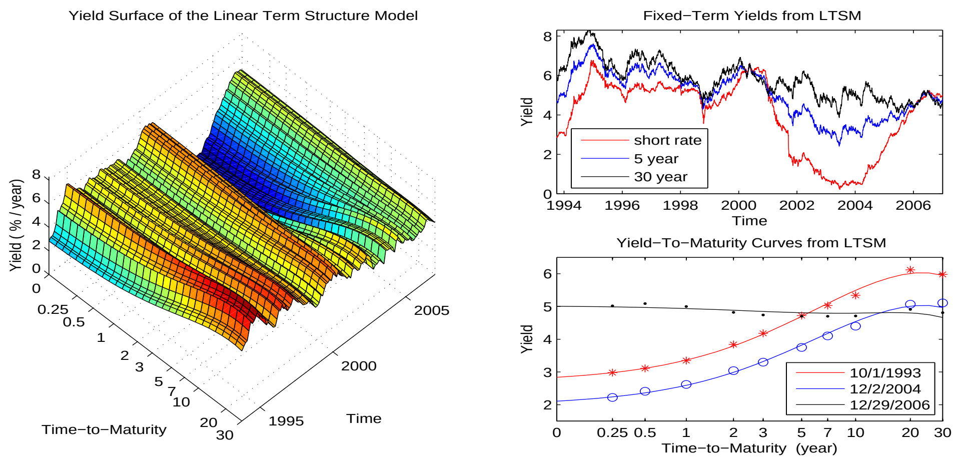
Critical information on US government debt is made public by law. Maximizing the number of bonds with available historical data, we found the data from US Department of the Treasury. We use 3313 complete sets of daily data from 10/1/1993 to 12/29/2006 for 10 different bonds with time-to-maturities 3-month, 6-month, 1-

year, 3-year, 5-year, 7-year, 10-year, 20-year, and 30-year, respectively. The data was computed or interpolated to be the zero-coupon bond rates.

Figure 2: Fixed-Term Yield-to-Maturity of US Government Bonds



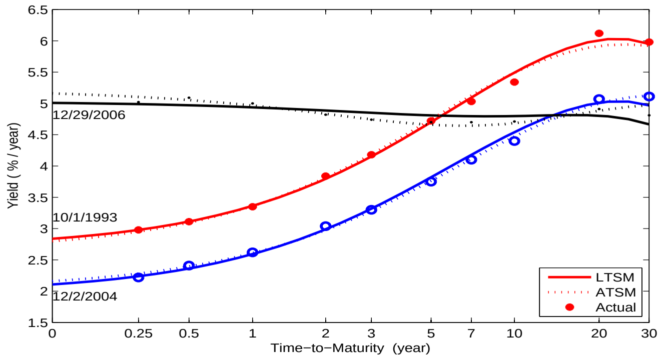
From the Original Data



From the Linear Term Structure Model

The fixed-term bond yields on every first trading day of October are listed in Table 1, where the last four rows are statistics of daily data covering the time period. Empirical densities of yields of bonds of different terms are plotted in Figure 1. It is quite clear that the fixed-term yields are not normally distributed: each empirical kurtosis is well-below 3 (that of normal distribution); the short-term yields have statistically significant negative skewness whereas long-term yields have positive skewness.

Figure 3: Comparison of Yield Curves from Original Data, ATSM, and LTSM.



The top part of Figure 2 illustrates the dynamical (in time t for fixed time-to-maturity $s = T - t$) and the cross–sectional (in s for fixed t), as well the overall (in (t, s)) behavior of the yields y_s^{t+s}. For readability, we only illustrate three historical yield curves (for fixed terms) and three yield-to-maturity curves (for fixed dates). All the others are similar and may be read from the yield surface. The yield-to-maturity curve at any current time t provides a window for an outlook of future economy. In general long term rates are higher than short term rates, but occasionally the reverse occurs, which can be significant to an economist. The dynamical behavior provides valuable information for statistical investigation of factors.

A fixed-term-yield y_t^{t+s} gives a bond price $Z_t^{t+s} = \exp(-s y_s^{t+s})$. It is in general very hard to gather historical data $\{Z_t^T\}_{t \leqslant T}$ for any fixed T. Of course, we can interpolate bond prices for fixe maturity date from empirical discrete data. Below we shall investigate empirical formulations derived from the affine term structure model (ATSM) and its general version (LTSM). The lower half of Figure 2 illustrates the reproduction of yields from our empirical LTSM model, where "$\cdot\ \circ\ *$ " represent empirical data.

2.5.2 Summary

Our empirical analysis consists of the following:

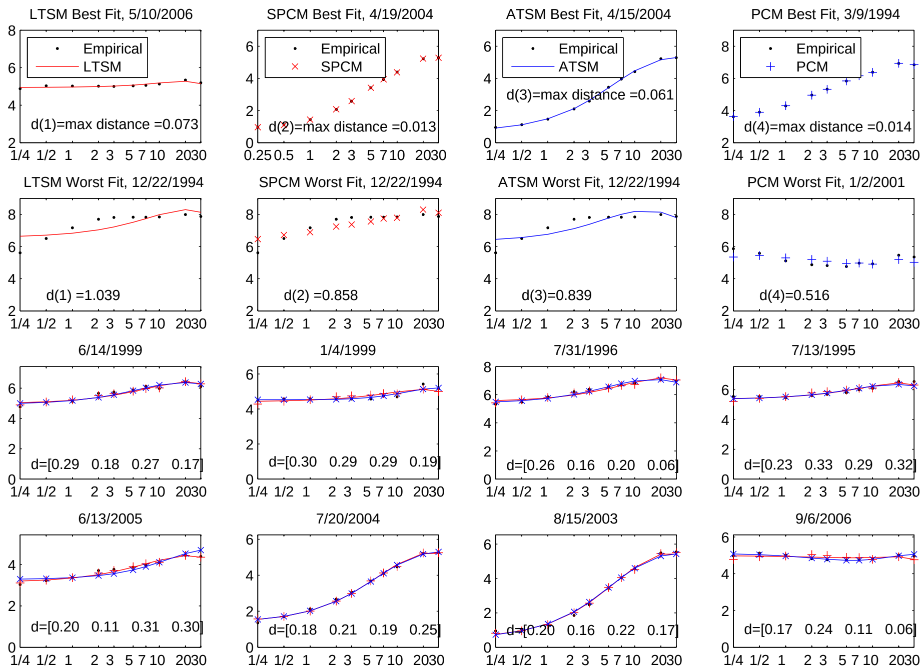
1. First we use the principal component analysis in §2.4 finding loads $a_k(s), \ell_k(s)$ and factors X_t^k and F_t^k in the Principal Component and General Principal Component models:

$$y_t^{t+s} = a_0(s) + a_1(s)X_t^1 + a_2(s)X_t^2 \qquad \forall\, t \in \mathbf{T},\ s \in \mathbf{S}, \tag{PCM}$$

$$y_t^{t+s} = \ell_1(s)F_t^1 + \ell_2(s)F_t^2 \quad \forall\, t \in \mathbf{T},\ s \in \mathbf{S}, \tag{SPCM}$$

where $\mathbf{T} = \{t_i\}_{i=1}^{3313}$ is the historical trading dates and $\mathbf{S} = \{1/4, 1/2, 1, 2, 3, 5, 7, 10, 20, 30\}$ is the list of terms of bonds under investigation.

Figure 4: Fitness of PCM, SPCM, ATSM and LTSM



2. Next we use linear regression to find constants $\{\sigma_k^{ij}\}$ in the following expressions:

$$\begin{aligned}\frac{\Delta X_t^i \Delta X_t^j}{\Delta t} &= \sigma_0^{ij} + \sigma_1^{ij} X_t^1 + \sigma_2^{ij} X_t^2 + \text{white noise},\\ \frac{\Delta F_t^i \Delta F_t^j}{\Delta t} &= \sigma_1^{ij} F_t^1 + \sigma_2^{ij} F_t^2 + \text{white noise}\end{aligned}$$

where $\Delta X_t^k = X_{t+\Delta t}^k - X_t^k$, $\Delta F_t^k = F_{t+\Delta t}^k - F_t^k$, and $\Delta t = t_{i+1} - t_i$ is one trading day. Here σ_k^{ij} in ATSM is different from that in LTSM; using same σ_k^{ij} is for notational simplicity.

Table 2: Effectiveness of Principal Component Models

(a) Individual and Cumulative Contributions of Factors

Factors k	1	2	3	4	5	6	7	8	9	10
					SPCM Norm^2					
λ_k	261.88	3.7994	0.1838	0.0137	0.0055	0.0033	0.0012	0.0009	0.0006	0.0003
Individual	0.9849	0.0143	0.0007	5.2e-5	2.1e-5	1.2e-5	5. e-6	3. e-6	2. e-6	1. e-6
Cumulative	0.9849	0.9992	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1
Remainder	0.0151	0.0008	0.0001	4.4 e-5	2.4e-5	1.1e-5	7. e-6	3. e-6	1. e-6	0
					SPCM Variances					
Individual	0.7943	0.1947	0.0096	0.0007	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000
Cumulative	0.7943	0.9891	0.9987	0.9994	0.9997	0.9998	0.9999	1.0000	1.0000	1
Remainder	0.2057	0.0109	0.0013	0.0006	0.0003	0.0002	0.0001	0.0000	0.0000	0
Factors	mean	1	2	3	4	5	6	7	8	9
					PCM Norm^2					
Individual	0.9289	0.0648	0.0059	0.0003	5.0e-5	1.9e-5	1.1e-5	4. e-6	3. e-6	2. e-6
Cumulative	0.9289	0.9937	0.9996	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Remainder		0.0711	0.0063	0.0004	9.0e-5	4.0e-5	1.1e-5	6. e-6	3. e-6	1. e-6
					PCM Variances					
$\hat{\lambda}_k$		17.217	1.5736	0.0857	0.0133	0.0050	0.0029	0.0011	0.0009	0.0005
individual		0.9109	0.0833	0.0045	0.0007	0.0003	0.0002	0.0001	0.0000	0.0000
Cumulative		0.9109	0.9942	0.9987	0.9994	0.9997	0.9998	0.9999	1.0000	1.0000
Remainder		0.0891	0.0058	0.0013	0.0006	0.0003	0.0002	0.0001	0.0000	0.0000

3. Supplied with known values of $\{\sigma_k^{ij}\}$ we solve the Riccati equations to obtain

$$y_t^{t+s} = \frac{A_0(s)}{s} + \frac{A_1(s)}{s} X_t^1 + \frac{A_2(s)}{s} X_t^2 \quad \forall\, t \in \mathbb{R}, s \in [0, T^{\max}), \qquad \text{(ATSM)}$$

$$y_t^{t+s} = \frac{L_1(s)}{s} F_t^1 + \frac{L_2(s)}{s} F_t^2 \quad \forall\, t \in \mathbb{R}, s \in [0, T^{\max}). \tag{LTSM}$$

The coefficients $\{r_k, p_k^i\}$ in the Riccati equations are chosen in a way such that the resulting loadings $A_i(s)/s$ and $L_i(s)/s$ match the best on **S** to those $a_i(s)$ in PCM and $\ell_i(s)$ in SPCM, respectively.

(b) Relative Size of Remainders

n	Term	1/4	1/2	1	2	3	5	7	10	20	30	Overall
						η_n in n-SPCM						
1	Var	0.268	0.237	0.175	0.096	0.046	0.017	0.066	0.205	0.691	0.887	0.206
	Std	0.518	0.487	0.419	0.310	0.214	0.132	0.257	0.453	0.831	0.942	0.454
	Mean	0.032	0.032	0.029	0.022	0.014	0.001	0.008	0.015	0.026	0.028	0.002
2	Var	0.019	0.006	0.002	0.010	0.014	0.013	0.007	0.003	0.018	0.031	0.011
	Std	0.139	0.078	0.041	0.100	0.120	0.115	0.085	0.053	0.135	0.175	0.105
	Mean	0.006	0.004	0.000	0.004	0.004	0.004	0.002	0.001	0.003	0.003	0.000
3	Var	0.002	0.001	0.002	0.001	0.000	0.001	0.002	0.002	0.003	0.004	0.001
	Std	0.039	0.025	0.040	0.025	0.017	0.033	0.048	0.045	0.053	0.064	0.037
	Mean	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	Var	0.000	0.000	0.000	0.001	0.000	0.001	0.001	0.002	0.003	0.001	0.001
	Std	0.015	0.022	0.016	0.022	0.017	0.024	0.023	0.045	0.053	0.037	0.025
	Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
						ε_n in n-PCM						
1	Var	0.101	0.072	0.031	0.007	0.010	0.054	0.105	0.189	0.348	0.430	0.089
	Std	0.318	0.268	0.175	0.083	0.099	0.231	0.324	0.435	0.590	0.656	0.298
2	Var	0.010	0.002	0.002	0.006	0.006	0.004	0.003	0.002	0.013	0.027	0.006
	Std	0.097	0.042	0.042	0.075	0.079	0.066	0.053	0.046	0.115	0.164	0.076
3	Var	0.001	0.001	0.002	0.001	0.000	0.001	0.002	0.002	0.003	0.004	0.001
	Std	0.035	0.023	0.040	0.023	0.017	0.034	0.048	0.044	0.051	0.066	0.036
4	Var	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.001	0.001
	Std	0.014	0.021	0.015	0.021	0.017	0.025	0.022	0.044	0.050	0.031	0.024
						the Difference Between n-PCM and n-SPCM						
2	Norm	0.036	0.023	0.002	0.022	0.026	0.021	0.012	0.006	0.0174	0.0189	0.020
	Var	0.008	0.004	0.000	0.005	0.008	0.008	0.004	0.001	0.012	0.015	0.006
	std	0.091	0.061	0.006	0.067	0.088	0.087	0.059	0.034	0.112	0.125	0.070
3	Norm	0.005	0.004	0.001	0.002	0.001	0.002	0.002	0.002	0.003	0.001	0.002
	Var	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.001	0.001
	Std	0.014	0.022	0.016	0.022	0.017	0.024	0.023	0.045	0.053	0.037	0.025

"Term" refers to time-to-maturity, "Var" sample variance, and "Std" sample standard deviation.

Figure 2 (lower half) illustrates the yields from our 2-factor LTSM. The result from ATSM is similar. Some of the tiny differences between ATSM, LTSM, and original data can be seen from the three yield curves in Figure 3. The overall fit of the four models: PCM, SPCM, ATSM, and LTSM, to the original data is illustrated by 16 plots in Figure 4, where the dates for the last 8 plots are "randomly" picked.

More details are carried out in the subsequent subsections.

2.5.3 Principal Component Analysis

According to the theory in §2.4.1 and its implementation in §2.4.2, we decompose all historical fixed-term-yields into principal components:

$$y_t^{t+s} - a_0(s) = \sum_{i=1}^{m} X_t^i\, a_i(s) \qquad \forall\, t \in \mathbf{T}, s \in \mathbf{S} \quad \Big(a_0(s) := \frac{1}{|\mathbf{T}|}\sum_{t\in\mathbf{T}} y_t^{t+s}\Big), \tag{2.5.1}$$

$$y_s^{t+s} = \sum_{i=1}^{m} F_t^i\, \ell_i(s) \qquad \forall\, t \in \mathbf{T}, s \in \mathbf{S}\,. \tag{2.5.2}$$

Here the principal components $\{F_t^k\}, \{X_t^i\}$ and loads $\{\ell_k\}, \{a_i\}$ are normalized such that

$$\frac{1}{|\mathbf{T}|}\sum_{t\in\mathbf{T}} F_t^k F_t^l = \delta^{kl}, \qquad \frac{1}{|\mathbf{T}|}\sum_{t\in\mathbf{T}} X_t^k X_t^l = \delta^{kl} \quad \forall\, k,l = 1,\cdots,m,$$
$$\sum_{s\in\mathbf{S}} a_i(s)a_j(s) = \lambda_i\delta_{ij}, \qquad \sum_{s\in\mathbf{S}} \ell_i(s)\ell_j(s) = \bar\lambda_i\delta_{ij} \quad \forall\, i,j = 1,\cdots,m.$$

The numerical procedure goes as follows.

> Let $Y = (y_{t_i}^{t_i+s_j})_{N\times m}$ be the **yield matrix**. Let $\Lambda = \mathrm{diag}(\lambda_1,\cdots,\lambda_m)$ $(\lambda_1 \geqslant \cdots \geqslant \lambda_m)$ and E be matrices such that $Y^\top Y = E\Lambda E^\top, EE^\top = \mathbf{I}_{m\times m}$. Setting $(F_{t_i}(s^j))_{T\times m} = YE\Lambda^{-1/2}$ and $(\ell_i(s^j))_{m\times m} = \Lambda^{1/2}E^\top$ we then obtain (2.5.2) with required normalization.
> Similarly, working on $\hat Y = (y_{t_i}^{t_i+s_j} - a_0(s_j))_{N\times m}$ we obtain (2.5.1).

Define $\|y\|^2 = \sum_{t\in\mathbf{T},s\in\mathbf{S}} |y_t^{t+s}|^2$ and $\mathrm{Var}[y] = \sum_{s\in\mathbf{S}} \mathrm{Var}[y_\cdot^{\cdot+s}]$. Then there are the Pythagorean identities

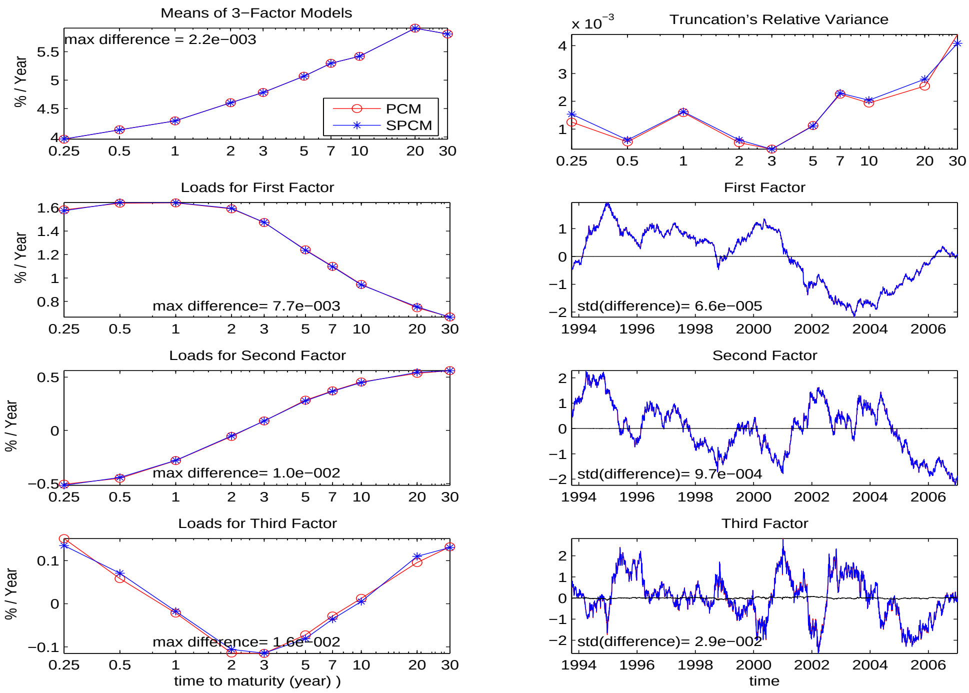
$$\|y\|^2 = \sum_{k=1}^{m} \|F_\cdot^k\, \ell_k(\cdot)\|^2 = \sum_{k=1}^{m} \lambda_k, \qquad \mathrm{Var}[y] = \sum_{k=1}^{m} \mathrm{Var}[X_\cdot^k a_k(\cdot)] = \sum_{k=1}^{m} \bar\lambda_k.$$

For each $k = 1,\cdots,m$, the individual contributions of $F_t^k\ell_k(s), X_t^k a_k(s)$ and cumulative contributions of $\sum_{i=1}^k F_t^i\ell_i(s)$, $a_0 + \sum_{i=1}^k X_t^i a_i(s)$ towards the total are listed in Table 2 (a).

To see the effect of the truncation of principal components, we write

$$y_t^{t+s} = \sum_{k=1}^{n} F_t^k \, \ell_k(s) + \eta_n(t,s) = a_0(s) + \sum_{k=1}^{n} X_t^n \, a_k(s) + \varepsilon_n(t,s) \quad \forall\, t \in \mathbf{T}, s \in \mathbf{S}.$$

Figure 5: Comparison of 3 factor PCM and SPCM)



The differences between the principal components of PCM and the mean deducted rotated principal components of SPCM are the middle curves in the figures, with standard deviation stated.

In Table 2 (b) we list the relative sizes of truncations $\eta_n(\cdot, s)$ and $\varepsilon_n(\cdot, s)$, as well as the difference between n-PCM and n-SPCM, $[\sum_{k=1}^{n} F_t^k \;\; \ell_k(s)] - [a_0(s) + \sum_{k=1}^{n} X_t^n \, a_k(s)]$, for different number n of principal components and time-to-maturity s (in year).

Part of Figure 4 displays the yield-to-maturity curves for the original data, $\{y_y^{t+s}\}_{s\in\mathbf{S}}$, the 2-PCM, $a_0(s) + X_t^1 a_1(s) + X_t^2 a_s(s)$, and the 2-SPCM, $F_t^1 \ell_1(s) + F_t^2 \ell_2(s)$, for a number of randomly picked dates, together with the dates of best and worst fits.

Remark 2.5.1. (1) It is quite clear from Table 2 and Figure 4 that principal 2-component models balance simplicity and accuracy. Indeed, the 2-PCM, $a_0(s) + a_1(s)X_t^1 + a_2(s)X_t^2$, contains about 99.96% of the total square norm and 99.42% of total variance, whereas the 2-SPCM, $\ell_1(s)F_t^1 + \ell_2(s)F_t^2$, contains about 99.92% of total square norm and 98.91% of total variances. Also SPCM is extremely close to PCM.

(2) The original data contains only two decimal points in percentage quotation. As the average size of the data is about 5%, the original data itself contains a relative truncation error of size about 0.005%/5% = 0.001. Hence, except for the first four principal components, the remaining ones should not be taken seriously. This fact may also be seen from the speed of decrease of the sizes of eigenvalues λ_k (or $\bar{\lambda}_k$) in k: it is extremely fast when $k \leqslant 3$ and relatively slow when $k \geqslant 4$.

2.5.4 Factor Rotation

To compare detailed structures of PCM and SPCM, we take appropriate bases of the principal subspace; this procedure is known as **factor rotation**.

In vector notation, we set $F_t = (F_t^1 \ \cdots \ F_t^n)$ and $\ell = (\ell_i(s^j))_{n\times m}$. Now let Q be an orthogonal matrix: $Q^\top Q = QQ^\top = \mathbf{I}_{n\times n}$ and set $\tilde{F}_t = F_t Q$ and $\tilde{\ell} = Q^\top \ell$. Then

$$\begin{aligned} Y_t^{\text{n}-\text{SPCM}} &:= \left(\sum_{k=1}^n F_t^k \,\ell_k(s^1), \ \cdots \ , \sum_{k=1}^n F_t^k \ell_k(s^m)\right) = F_t \ell \\ &= (F_t Q)(Q^\top L) = \tilde{F}_t \,\tilde{\ell} = \ell_0^n(s) + \sum_{k=1}^n \left(\tilde{F}_t^k - \tilde{m}_k\right) \tilde{\ell}_k(s) \end{aligned}$$

where

$$\tilde{m}^k := \frac{1}{|\mathbf{T}|}\sum_{t=\mathbf{T}}^n \tilde{F}_t^k, \qquad \ell_0^n(s) := \frac{1}{|\mathbf{T}|}\sum_{t\in\mathbf{T}} Y_t^{\text{n}-\text{SPCM}}(s) = \sum_{t\in\mathbf{T}} \tilde{m}^k \tilde{\ell}_k(s).$$

We choose the best Q such that $\sum_{k=1}^n \mathrm{Var}[X^k - \tilde{F}^k]$ is minimized. For $n = 3$, the resulting factors $\{X_t^k\}$ and loads $a_k(\cdot)$ for the PCM and mean-deducted factors $\{\tilde{F}_t^k -$

$\tilde{m}^k\}$ and loads $\tilde{\ell}_k(\cdot)$ of the SPCM are plotted in Figure 5, from which we see that the two models are in perfect agreement. The case $n = 2$ is similar and hence is omitted.

Remark 2.5.2. Referring to Figure 5, the load $a_1(\cdot)$ of the first factor is uniformly positive, which means an increment of the first factor X_t^1 increases the yield y_t^T for every $T = t + s > t$. Thus, the first factor is commonly referred to as the **level factor**. The load $a_2(\cdot)$ of the second factor is monotonic, which means an increment of the second factor X_t^2 rotates the yield-to-maturity curve, so the second factor is called the **slope factor**. Similarly, the third factor is called the **curvature factor**; see [81].

In choosing bases for a principal subspace, we prefer to use the principal components since under this base the contribution of the term $X_t^k a_k(s)$ is proportional to $\bar{\lambda}_k$ which decreases in k rapidly; indeed our numerical calculation confirms this choice. Thus, in the sequel, for the PCM, we take the principal components (X_t^1, X_t^2) as our working factors. For SPCM, we make a simple rotation according to the following

$$F_t^1\ell_1(s) + F_t^2\ell_2(s) = \frac{F_t^1}{c_1}\, c_1[\ell_1(s) + c\ell_2(s)] + \frac{[F_t^2 - cF_t^1]}{c_2}\, c_2\ell_2(s).$$

We take $\tilde{F}_t = F_t^1/c_1$, $\tilde{\ell}_1(s) = c_1[\ell_1(s) + c\ell_2(s)]$, $\tilde{F}_t^2(s) = [F_t^2 - cF_t^1]/c_2$, $\tilde{\ell}_2 = c_2\ell_2(s)$ where

$$c = \frac{\text{mean}(F^2)}{\text{mean}(F^1)}, \qquad c_1 = \text{mean}(F_1), \qquad c_2 = \text{std}(F^2 - cF^1).$$

where mean is the sample mean and *std* is sample standard deviation. The rotated factors satisfy $\text{mean}(\tilde{F}^1) = 1, \quad \text{mean}(\tilde{F}^2) = 0, \quad \text{std}(\tilde{F}^2) = 1$. Since c is small, the term $\tilde{F}_t^2\tilde{\ell}_2(s)$ is as small as $F_t^2\ell_2(s)$. Hence, this new base can still be regards as principal components. In the sequel, we drop the $\tilde{}$ sign.

2.5.5 Empirical Two Factor Models

In summary, we obtain two empirical models:

$$\begin{pmatrix} y_t^{t+1/4} \\ y_t^{t+1/2} \\ y_t^{t+1} \\ y_t^{t+2} \\ y_t^{t+3} \\ y_t^{t+5} \\ y_t^{t+7} \\ y_t^{t+10} \\ y_t^{t+20} \\ y_t^{t+30} \end{pmatrix} = \begin{pmatrix} 3.94 \\ 4.11 \\ 4.28 \\ 4.62 \\ 4.80 \\ 5.09 \\ 5.31 \\ 5.42 \\ 5.90 \\ 5.80 \end{pmatrix} F_t^1 + \begin{pmatrix} 0.86 \\ 0.84 \\ 0.72 \\ 0.49 \\ 0.27 \\ \text{-}0.08 \\ \text{-}0.29 \\ \text{-}0.49 \\ \text{-}0.79 \\ \text{-}0.85 \end{pmatrix} F_t^2, \tag{2-SPCM}$$

Here the factors $X^i = \{X_t^i\}_{t\in\mathbf{T}}, F^i = \{F_t^i\}_{t\in\mathbf{T}}$ are normalized so that, in sample statistics,

$$\mathrm{Cov}(X^i, X^j) = \delta_{ij}, \quad \mathrm{mean}(X^i) = 0, \quad \mathrm{mean}(F^1) = 1, \quad \mathrm{mean}(F^2) = 0, \quad \mathrm{std}(F^2) = 1.$$

We find that $\mathrm{std}(F^1) = 0.247$. The correlation $\mathrm{cor}(F^1, F^2) = 0.75$ is not small, but it should not be a concern since F_t^1 is close to the constant 1. In part of Figure 4 we illustrated the accuracy of the 2-PCM and 2-SPCM. From the figure and also Table 2 one concludes that the two factor models fit the historical data very well.

Remark 2.5.3. (1) The 2-SPCM can be written as

$$y_t^{t+s} = \ell_1(s) + \ell_1(s)(F_t^1 - 1) + \ell_2(s)F_t^2 \quad \forall\, t \in \mathbf{T}, s \in \mathbf{S},$$

where ℓ_1 and ℓ_2 are the coefficient columns of F_t^1 and F_t^2 in 2-SPCM, respectively. Here the load $\ell_1(\cdot)$ serves as both the mean and the load of the first factor. In this formulation, we can read the mean of the model from the load of the first factor; this is indeed the reason that we make the special rotation from the original principal component formulation. It is clear that our 2-SPCM is simpler than the traditional 2-PCM, whereas both have similar accuracy. As mentioned, since a_0 represents the mean of yields and in general it is very hard to measure empirically the mean of return of financial securities, making a_0 disappear as in 2-SPCM becomes very attractive.

$$\begin{pmatrix} y_t^{t+1/4} \\ y_t^{t+1/2} \\ y_t^{t+1} \\ y_t^{t+2} \\ y_t^{t+3} \\ y_t^{t+5} \\ y_t^{t+7} \\ y_t^{t+10} \\ y_t^{t+20} \\ y_t^{t+30} \end{pmatrix} = \begin{pmatrix} 3.96 \\ 4.13 \\ 4.28 \\ 4.60 \\ 4.78 \\ 5.07 \\ 5.30 \\ 5.42 \\ 5.92 \\ 5.81 \end{pmatrix} + \begin{pmatrix} 1.58 \\ 1.64 \\ 1.64 \\ 1.59 \\ 1.47 \\ 1.24 \\ 1.10 \\ 0.95 \\ 0.75 \\ 0.67 \end{pmatrix} X_t^1 + \begin{pmatrix} \text{-0.51} \\ \text{-0.45} \\ \text{-0.28} \\ \text{-0.06} \\ 0.09 \\ 0.28 \\ 0.37 \\ 0.45 \\ 0.54 \\ 0.56 \end{pmatrix} X_t^2. \tag{2-PCM}$$

(2) As shown in Tables 2, the relative size of variance of the truncation of the two factor model is about 1%. From Table 2, we see the behavior of relative variances of the 2-factor truncation in term of the time-to-maturity. The relative large 0.95% residual of 3-month bond and the 2.68% residual of the 30-year bond at both ends indicate that adding a curvature factor to the two factor model can reduce significantly the truncation error. Nevertheless, judging the balance between simplicity and accuracy, we conclude that at this stage the contribution of the curvature factor $X_t^3 a_3(s)$ or $F_t^3 \ell_3(s)$ is not significant enough to include in our model; see Table 2 and Figure 5 for the size of load of the third factor.

2.5.6 The Empirical Covariance Matrix

In 2-factor models, $\{(F_t^1, F_t^2)\}_{t\in\mathbb{R}}$ in LTSM and $\{(X_t^1, X_t^2)\}_{t\in\mathbb{R}}$ in ATSM are assumed to be Itô processes. The covariance matrix is defined by

$$\sigma_t := \frac{\mathrm{Cov}(dF_t^1, dF_t^2)}{dt} = \sum_{k=1}^{n} \sigma_k F_t^k, \tag{2.5.3}$$

$$\sigma_t := \frac{\mathrm{Cov}(dX_t^1, dX_t^2)}{dt} = \sigma_0 + \sum_{k=1}^{n} \sigma_k X_t^k \tag{2.5.4}$$

where $\sigma_0, \sigma_1, \sigma_2$ are constant 2×2 matrices; they have different values in the two different models. Here we explain our numerical procedure to estimate these matrices.

2.5.6.1 The Covariance Matrices in LTSM From the empirical sample path of the factors $\{F_t^k\}$, we can calculate the sample covariance matrix via

$$S_t = \frac{1}{\Delta t}\begin{pmatrix} \Delta F_t^1 \Delta F_t^1 & \Delta F_t^1 \Delta F_t^2 \\ \Delta F_t^2 \Delta F_t^1 & \Delta F_t^2 \Delta F_t^2 \end{pmatrix}, \qquad \Delta F_t^k = F_{t+\Delta t}^k - F_t^k, \quad t \in \tilde{\mathbf{T}} = \{t_i\}_{i=1}^{N-1}.$$

We can express S_t as a linear combination of F_t^1 and F_t^2:

$$S_t = \sigma_1 F_t^1 + \sigma_2 F_t^2 + \text{white noise}. \tag{2.5.5}$$

We propose three methods to find the matrices σ_1 and σ_2.

1. Linear Regression for the Original. We use linear regression to find estimator

$$\begin{aligned}(\sigma_1^{(\mathrm{I})}, \sigma_2^{(\mathrm{I})}) &= \operatorname*{argmin}_{(\tilde\sigma_1, \tilde\sigma_2)} \sum_{i=1}^{N-1} \| S_{t_i} - \tilde\sigma_1 F_{t_i}^1 - \tilde\sigma_2 F_{t_i}^2 \|^2 \\ &= \left(\begin{pmatrix} 0.024 & 0.000 \\ 0.000 & 0.177 \end{pmatrix}, \quad \begin{pmatrix} 0.009 & 0.008 \\ 0.008 & 0.025 \end{pmatrix} \right).\end{aligned}$$

Since sample mean$(F_t^1, F_t^2) = (1, 0)$ we have estimation

$$\sigma_t = \sigma_1 + \sigma_1(F_t^1 - 1) + \sigma_2 F_2^t \approx \begin{pmatrix} 0.024 & 0.000 \\ 0.000 & 0.177 \end{pmatrix} + \sigma_1^{(I)}(F_t^1 - 1) + \sigma_2^{(I)} F_t^2.$$

Since sample std$(F^1, F^2) = (0.25, 1)$, we see that σ_t is not far away from a constant matrix.

2. Linear Regression for the Integral. We integrate (2.5.5) with respect to t to obtain

$$\int S_t dt = \sigma_1 \int F_t^1 dt + \sigma_2 \int F_t^2 dt + \text{white noise}.$$

Using linear regression on this relation we find

$$\begin{aligned} (\sigma_1^{(\mathrm{II})}, \sigma_2^{(\mathrm{II})}) &= \underset{(\tilde{\sigma}^1, \tilde{\sigma}^2)}{\operatorname{argmin}} \min_c \sum_{i=1}^{N-1} \left\| \sum_{j=1}^{i} \left\{ S_{t_j} - \tilde{\sigma}_1 F_{t_j}^1 - \tilde{\sigma}_2 F_{t_j}^2 \right\} - c \right\|^2 \\ &= \left(\begin{pmatrix} 0.025 & 0.000 \\ 0.000 & 0.203 \end{pmatrix}, \quad \begin{pmatrix} 0.008 & 0.009 \\ 0.009 & 0.052 \end{pmatrix} \right). \end{aligned}$$

3. Linear Regression for the Moving Average. We write (2.5.5) in moving average:

$$\frac{1}{h} \int_{t-h}^{t} S_t dt = \frac{\sigma_1}{h} \int_{t-h}^{t} F_t^1 dt + \frac{\sigma_2}{h} \int_{t-h}^{t} F_t^2 dt + \text{white noise}$$

Taking $h = H\Delta t = 0.4$ (year) and apply linear regression we obtain

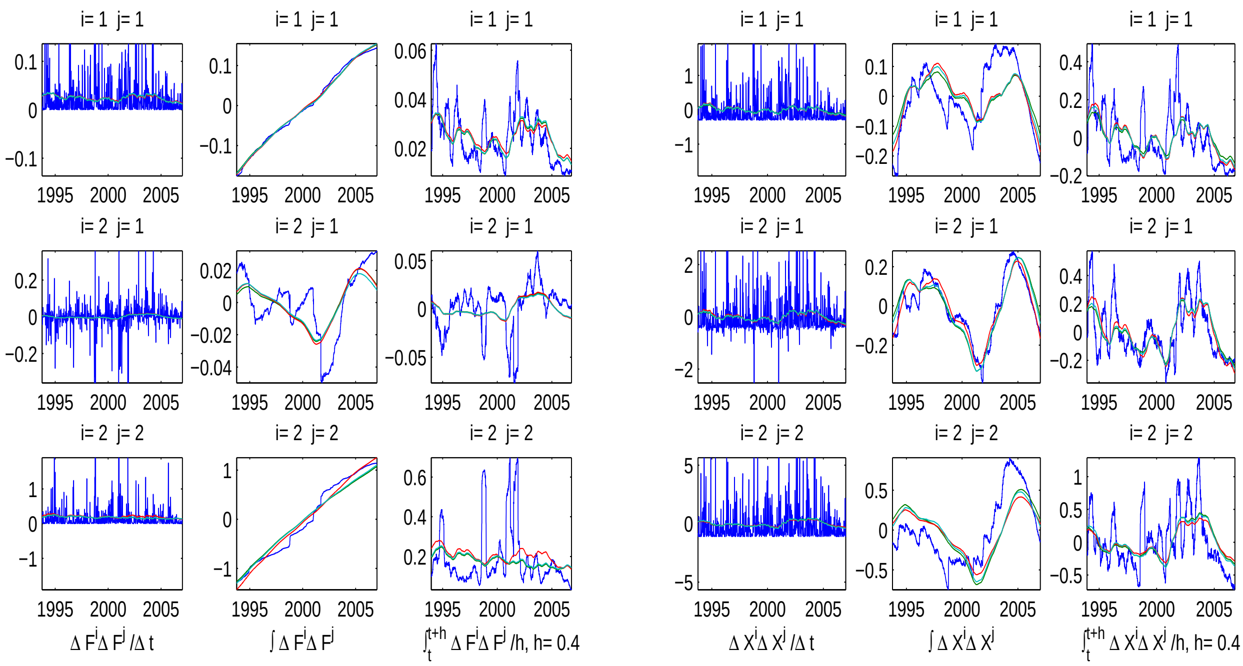
$$\begin{aligned} (\sigma_1^{(\mathrm{III})}, \sigma_2^{(\mathrm{III})}) &= \underset{(\tilde{\sigma}^1, \tilde{\sigma}^2)}{\operatorname{argmin}} \sum_{i=1}^{N-H-1} \left\| \sum_{j=i}^{i+H} \left\{ S_{t_j} - \tilde{\sigma}_1 F_{t_j}^1 - \tilde{\sigma}_2 F_{t_j}^2 \right\} \right\|^2 \\ &= \left(\begin{pmatrix} 0.025 & 0.000 \\ 0.000 & 0.181 \end{pmatrix}, \quad \begin{pmatrix} 0.009 & 0.008 \\ 0.008 & 0.026 \end{pmatrix} \right). \end{aligned}$$

For each approximation of (σ_1, σ_2), we compute $\sigma_1 F_t^1 + \sigma_2 F_t^2, \sigma_1 \int F_t^1 + \sigma_2 \int F_t^2$, and $h^{-1}\sigma_1 \int_t^{t+h} F_s ds + h^{-1}\sigma_2 \int_t^{t+h} F_s^2 ds$ and compare them with the corresponding S_t, $\int S_t$, and $h^{-1}\int_t^{t+h} S_s ds$. The results are displayed in Figure 6 (a). The linear regressions fits the target, not superb, but reasonably well.

At this moment, we would like to point out that there are a few number of increments ΔF which may be more appropriate to be characterized as "jumps". With $h = 0.4$ (year), the moving average still cannot smoothen these jumps; see the last column in Figure 6(a). This suggests that it maybe better to include jumps in the model.

Figure 6: Linear Regression for the Covariance Matrices

(a) SPCM (b) PCM



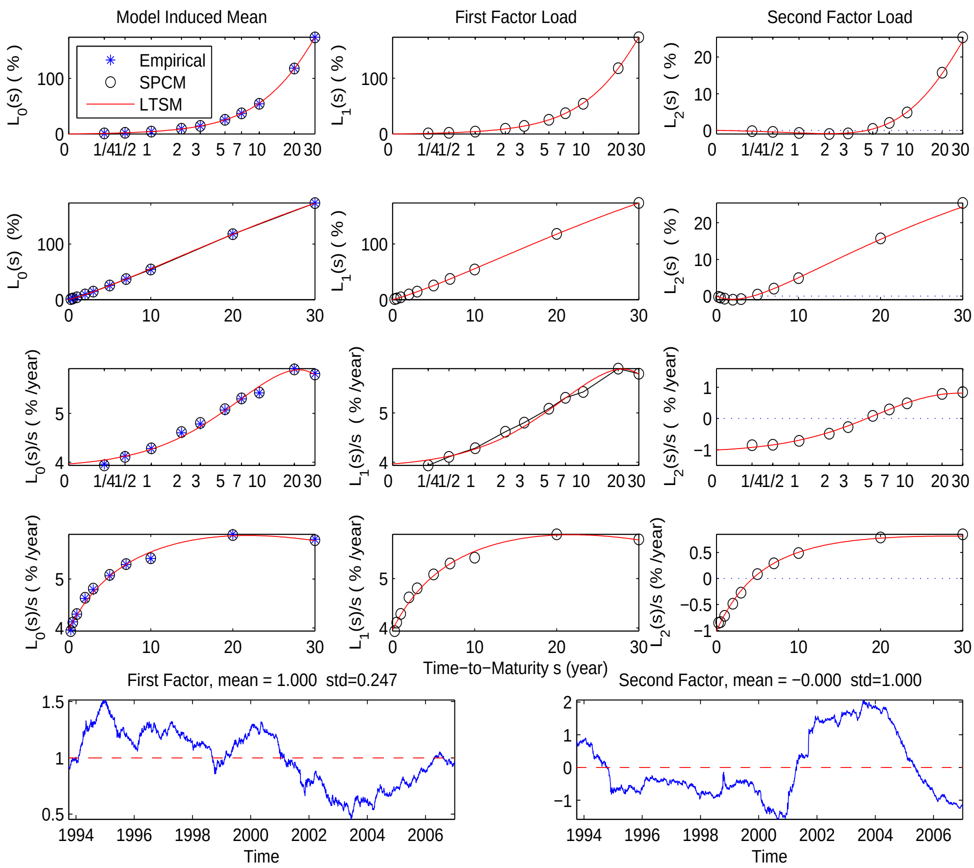
Smooth curves are linear regressions; non-smooth curves are from the targets.

The matrices σ_1 and σ_2 obtained from the three different methods agree to each other. The relative small size of std(F^1) and σ_2 indicates that σ_t has a small variance. From now on we fix σ_1 and σ_2 to be the averages of three estimators. Writing

$z = (z^1, z^2) = (x, y)$ we have

$$\sigma(x,y) = \sigma_1 x + \sigma_2 y = \begin{pmatrix} 0.0245 & 0.0002 \\ 0.0002 & 0.1871 \end{pmatrix} x + \begin{pmatrix} 0.0085 & 0.0084 \\ 0.0084 & 0.0345 \end{pmatrix} y \quad (2.5.6)$$

Figure 7: Fitness Between Empirical and Theoretical Loads of LTSM



The LTSM loads are obtained by solving the Riccati equations (2.2.6). The induced mean is obtained by the formula $L_0 = L_1 \text{mean}(F^1) + L_2 \text{mean}(F^2) = L_1$. SPCM loads are obtained from the General Principal Component Model.

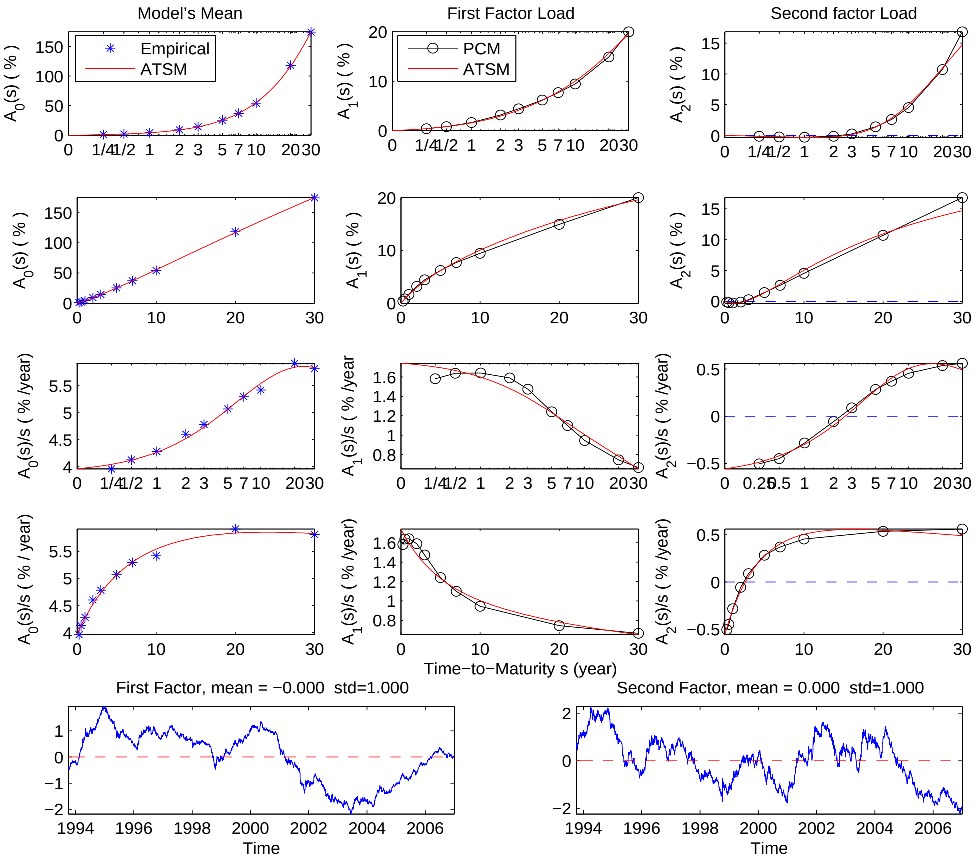
2.5.6.2 The Covariance Matrices in ATSM Using a similar linear regression procedure we can find estimators for $\sigma_0, \sigma_1, \sigma_2$ in (2.5.4) for the benchmark ATSM.

The average of the three estimators is as follows:

$$\sigma_0 = \begin{pmatrix} 0.30 & 0.40 \\ 0.40 & 1.11 \end{pmatrix}, \qquad \sigma_1 = \begin{pmatrix} 0.00 & -0.06 \\ -0.06 & -0.16 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0.08 & 0.14 \\ 0.14 & 0.17 \end{pmatrix}.$$

The fitness of the linear regressions to the targets is shown in Figure 6 (b).

Figure 8: Fitness Between Empirical and Theoretical Loads of ATSM



The ATSM loads are obtained by solving the Ricartti equation (2.2.6). The load $A_0(s)/s$ represents the mean yield. PCM loads are obtained from the Principal Components Model.

2.5.7 The Riccati Equations

Knowing the coefficient (σ_k^{ij}), in this subsection we solve the Riccati equations. We want to find parameters such that solutions of the Riccati equations for ATSM and SATM fit the best on $\mathbf{S}$ to that from the PCM and SPCM, respectively.

2.5.7.1 The Riccati 's Equations in LTSM First we consider the system of Ricartti equations (2.2.6), which can be written as

$$\frac{d\mathbf{L}}{ds} = P \begin{pmatrix} 1 \\ \mathbf{L} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \mathbf{L}^\top \sigma_1 \\ \mathbf{L}^\top \sigma_2 \end{pmatrix} \mathbf{L} \quad \forall\, s \in [0, s_m], \qquad \mathbf{L}|_{s=0} = (0,0)^\top. \tag{2.5.7}$$

where

$$\mathbf{L} = \begin{pmatrix} L_1(s) \\ L_2(s) \end{pmatrix}, \qquad P = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \end{pmatrix} := \begin{pmatrix} r_1 & p_1^1 & p_1^2 \\ r_2 & p_2^1 & p_2^2 \end{pmatrix}.$$

For simplicity, we assume that σ_1 and σ_2 are known 2×2 matrices. Given P, we denote the solution of (2.5.7) by $\mathbf{L}(P, s)$. Our purpose is to find a special P such that $\mathbf{L}(P, \cdot)$ matches empirical data via the SPCM, at the points in $\mathbf{S}$. For this, we introduce a $2 \times m$ matrix

$$\mathbf{L}(P, \mathbf{S}) = \Big(\mathbf{L}(P, s_1), \ \cdots \ , \ \mathbf{L}(P, s_m)\Big)_{2\times m}.$$

Also, we denote the empirical value of $\mathbf{L}$ at $\mathbf{S}$ by $\mathbf{L}^*$:

$$\mathbf{L}^*(s) = s\ell(s), \qquad \ell(s) := \begin{pmatrix} \ell_1(s) \\ \ell_2(s) \end{pmatrix} \qquad \forall\, s \in \mathbf{S}$$

where the values of $\ell_k(s)$, for $s \in \mathbf{S}$, are obtained from 2-SPCM. Hence, the problem here is to find the optimal P^* such that $\mathbf{L}(P^*, \mathbf{S})$ is as close to $\mathbf{L}^*(\mathbf{S}) := (\mathbf{L}^*(s_1), \cdots, \mathbf{L}^*(s_m))_{2\times m}$ as possible. More precisely, we want to solve the following minimization problem:

$$P^* := \operatorname*{argmin}_P \sum_{s\in\mathbf{S}}^{m} \omega^2(s)\left|\mathbf{L}(P,s)-\mathbf{L}^*(s)\right|^2, \qquad \omega(s) := \frac{1}{s}. \tag{2.5.8}$$

We solve this least square minimization problem in two steps.

Step 1: Initial Approximation. In this step, we want to find an initial guess of P^*. To do this, we integrate the ode in (2.5.7) over $[0, s]$ to obtain its equivalent integral formulation

$$\mathbf{L}(s) = P\,\mathcal{V}[\mathbf{L}(P,\cdot)](s) - [\mathbf{L}(P,\cdot)](s) \quad \forall\, s \in [0, s_m]$$

where $\mathcal{V}[\mathbf{L}]$ and $[\mathbf{L}]$ are operators defined by, for any give continuous $\mathbf{L} : [0, s_m] \to \mathbb{R}^2$,

$$\mathcal{V}[\mathbf{L}](s) = \begin{pmatrix} s \\ \\ \displaystyle\int_0^s \mathbf{L}(\tau)d\tau \end{pmatrix}, \qquad [\mathbf{L}](s) := \begin{pmatrix} \frac{1}{2}\displaystyle\int_0^s \mathbf{L}^\top(\tau)\sigma_1\mathbf{L}(\tau)\,d\tau \\ \frac{1}{2}\displaystyle\int_0^s \mathbf{L}^\top(\tau)\sigma_2\mathbf{L}(\tau)d\tau \end{pmatrix} \qquad \forall\, s \in [0, s_m].$$

Thus, our minimization problem (2.5.8) can be written as

$$\min \sum_{s\in\mathbf{S}} \omega^2(s)\Big|\mathbf{L}^*(s) + [\mathbf{L}(P,\cdot)](s) - P\mathcal{V}[\mathbf{L}(P,\cdot)](s)\Big|^2.$$

As an approximation, we replace ω by 1 and $\mathbf{L}(P,\cdot)$ by $\mathbf{L}^*(\cdot)$. Then the least square problem can be solved. The solution, denoted by $P^{(0)}$, is given by

$$P^{(0)} = \Big(\mathbf{L}^*(\mathbf{S}) + B^*(\mathbf{S})\Big)V^*(\mathbf{S})\Big(V^*(\mathbf{S})V^*(\mathbf{S})^\top\Big)^{-1}.$$

Here $V^*(s) := \mathcal{V}[\mathbf{L}^*](s)$ and $B^*(s) := [\mathbf{L}^*](s)$, for $s \in \mathbf{S}$, are numerically evaluated as follows: First we use the known values $\mathbf{L}^*$ on $\mathbf{S}$ to construct a cubic spline interpolation $\mathbf{L}^*(\cdot)$ on $[0, s_m]$; then we use the Simpson's quadrature rule to find the numerical approximation for the integral defining $\mathcal{V}[\mathbf{L}^*]$ and $[\mathbf{L}^*]$.

Using σ_1 and σ_2 obtained in the previous subsection and the column vectors in

2-SPCM as (ℓ_1, ℓ_2), we obtain the initial guess

$$P^{(0)} = \begin{pmatrix} 4.119 & 0.068 & -0.257 \\ -0.976 & 0.071 & -0.350 \end{pmatrix},$$

with error

$$\begin{aligned} \max_{k=1,2,s\in\mathbf{S}} |L_k(P^{(0)}, s)/s - \ell_k(s)| &= 0.25, \\ \Big(\tfrac{1}{2m}\textstyle\sum_{k=1}^2 \sum_{s\in\mathbf{S}} |L_k(P^{(0)}, s)/s - \ell_k(s)|^2\Big)^{1/2} &= 0.074. \end{aligned}$$

Step 2: Newton's Iteration. In this step, we use Newton's iteration. Note that

$$\mathbf{L}(P + \Delta P, s) = \mathbf{L}(P, s) + D_p\mathbf{L}(P, s)\ \Delta P + O(\|\Delta P\|^2)$$

where

$$D_p\mathbf{L}(P, s) = \Big(\frac{\partial \mathbf{L}}{\partial p_1}, \cdots, \frac{\partial \mathbf{L}}{\partial p_6}\Big)$$

is the variation of $\mathbf{L}$ with respect to $\mathbf{L}$. Knowing P, $D_p\mathbf{L}(P, s)$ can be obtained by solving the system of differential equations

$$\begin{aligned} \frac{d}{ds} D_p\mathbf{L}(P, s) &= \left\{ \begin{pmatrix} p_2 & p_3 \\ p_5 & p_6 \end{pmatrix} - \begin{pmatrix} \mathbf{L}'(P, s)\sigma_1 \\ \mathbf{L}'(P, s)\sigma_2 \end{pmatrix} \right\} D_p\mathbf{L} + \\ &\quad \begin{pmatrix} 1 & 0 & L_1(P, s) & 0 & L_2(P, s) & 0 \\ 0 & 1 & 0 & L_1(P, s) & 0 & L_2(P, s) \end{pmatrix} \quad \forall\, s \in [0, s_m], \\ D_p\mathbf{L}(P, 0) &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

This, together with the odes for $\mathbf{L}(P, s)$, forms a closed system of 14 ordinary differential equations, which can be easily handled by a MatLab ode solver.

Now suppose we already have a value P that is close to P^*. Then writing $P^* =$

$P + \Delta P$, we can reformulate our minimization problem (2.5.8) as

$$\min_{\Delta P} \sum_{s \in \mathbf{S}} \omega^2(s) \Big| \mathbf{L}^*(s) - \mathbf{L}(P,s) - D_p \mathbf{L}(P,s) \Delta P + O(|\Delta P|^2) \Big|^2.$$

Omitting $O(|\Delta P|^2)$, such a problem has a closed from solution given by $\Delta P = (D'D)^{-1} D' \varepsilon$ where

$$\varepsilon = \begin{pmatrix} \omega(s_1)[\mathbf{L}^*(s_1) - \mathbf{L}(P,s_1)] \\ \vdots \\ \omega(s_m)[\mathbf{L}^*(s_m) - \mathbf{L}(P,s_m)] \end{pmatrix}, \qquad D = \begin{pmatrix} \omega(s_1) D_p \mathbf{L}(P,s_1) \\ \vdots \\ \omega(s_m) D_p \mathbf{L}(P,s_m) \end{pmatrix}.$$

Hence, the Newton's iteration becomes $P^{new} = P^{old} + \Delta P$.

Now suppose we have a fixed point, i.e., $\Delta P = 0$. Then we obtain a local minimizer, since any infinitesimal change from P to $P + dP$ will not decrease the value of the target function.

In our numerical calculation, the Newton's iteration stops in 4 iterations, with tolerance set at 10^{-5}. The final fixed point is

$$P^* = \begin{pmatrix} 3.9646 & 0.0793 & -0.3456 \\ -1.0099 & 0.0754 & -0.3937 \end{pmatrix}$$

with intrinsic error

$$\begin{aligned} \max_{k=1,2, s \in \mathbf{S}} |L_k(P^*,s)/s - \ell_k(s)| &= 0.12, \\ \left(\tfrac{1}{2m} \textstyle\sum_{k=1}^{2} \sum_{s \in \mathbf{S}} |L_k(P^*,s)/s - \ell_k(s)|^2 \right)^{1/2} &= 0.053. \end{aligned}$$

Hence, we obtain the estimators for the affine functions, writing $z = (x,y)$,

$$\begin{pmatrix} R(x,y) \\ \\ P^1(x,y) \\ \\ P^2(x,y) \end{pmatrix} = \begin{pmatrix} r_1 x + r_2 y \\ \\ p_1^1 x + p_2^1 y \\ \\ p_1^2 x + p_2^2 y \end{pmatrix} = \begin{pmatrix} 3.9646\, x - 1.0099\, y \\ \\ 0.0793\, x + 0.0754\, y \\ \\ -0.3456\, x - 0.3937\, y \end{pmatrix}. \tag{2.5.9}$$

Here R is in the unit of %/year. Note that mean$(F^1) = 1$ and mean$(F^2) = 0$, so mean$(R) = 3.9646$ (%/year).

The fitness between the loads from empirical data via the SPCM and the loads from solutions of Riccati equations is displayed in Figure 7, where two time-to-maturity scales are used: one is the standard scale; the other is the $\log[1/4 + s]$ scale for $s \in [0, 30]$, where actual values of s are marked. Using both scales, we can see the overall fit of the theoretical LTSM to the empirical data via SPCM. In the figure, we plot both the loads $L_i(s)$, $i = 1, 2$, in the bond price formula $-\log Z_t^{t+s} = L_1(s)F_t^1 + L_2(s)F_t^2$ and the loads $\ell_i(s) = L_i(s)/s$ in the yield formula $y_t^{t+s} = \ell_1(s)F_t^1 + \ell_2(s)F_t^2$. The model induced mean is calculated from $\ell_0(s) = \ell_1(s)\text{mean}(F^1) + \ell_2(s)\text{mean}(F^2) = \ell_1(s)$.

Using the solution $(L_1(s), L_2(s))$ of the Riccati equations (2.2.6) and the (rotated) factors (F_t^1, F_t^2) of SPCM, we then obtain the complete description of the LTSM model. The resulting yield surface and sample yield curves are displayed in the lower part of Figure 2. Comparison with empirical data and other models are shown in Figure 4. The relative sizes of differences between the empirical yields and model yields are show in Table 3 (a).

2.5.7.2 The Affine Term Structure Model Following a procedure similar to the one described above for the LTSM model, we can find optimal parameter such that the solution of the Riccati equations (2.2.4) matches on $\mathbf{S}$ that obtained from 2-PCM. The optimal parameters we obtained translate to the following:

$$\begin{pmatrix} R_t \\ P_t^1 \\ P_t^2 \end{pmatrix} = \begin{pmatrix} 0.0396 & 0.0174 & -0.0056 \\ -0.3398 & -0.1545 & 0.2597 \\ -0.1411 & 0.0868 & -0.2387 \end{pmatrix} \begin{pmatrix} 1 \\ X_t^1 \\ X_t^2 \end{pmatrix}. \tag{2.5.10}$$

The intrinsic difference is

$$\begin{aligned} \max_{k=0,1,2,s\in\mathbf{S}} |A_k(P^*,s)/s - a_k(s)| &= 0.12, \\ \left(\frac{1}{3m}\sum_{k=0}^{2}\sum_{s\in\mathbf{S}} |A_k(P^*,s)/s - a_k(s)|^2\right)^{1/2} &= 0.056. \end{aligned}$$

The fit between theoretical loads obtained by solving (2.2.4) and those from 2-PCM is shown in Figure 8.

Using solutions $(A_1(s), A_2(s))$ of the Riccati equation (2.2.4) and the factors (X_t^1, X_t^2) obtained from 2-PCM, we obtain a complete description of the ATSM model. The resulting yield surface and sample yield curves are similar to those in the lower part of Figure 2. Comparisons with empirical data and other models are shown in Figures 3 and 4. The relative sizes of the difference between empirical yields and ATSM yields are listed in Table 3 (b).

Table 3: Effectiveness of 2-Factor ATSM and LTSM

Term	1/4	1/2	1	2	3	5	7	10	20	30	Overall
(a) Relative Size of Difference between Empirical and LTSM											
Norm^2	0.004	0.001	0.000	0.002	0.002	0.001	0.000	0.001	0.000	0.001	0.001
Norm	0.062	0.028	0.017	0.042	0.041	0.029	0.018	0.026	0.022	0.028	0.029
Var	0.023	0.005	0.002	0.014	0.018	0.014	0.007	0.002	0.019	0.034	0.012
Std	0.151	0.074	0.045	0.117	0.133	0.118	0.084	0.046	0.139	0.184	0.112
(b) Relative Size of Difference between Empirical and ATSM											
Norm^2	0.003	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001	0.001	0.001
Norm	0.051	0.018	0.019	0.039	0.032	0.017	0.014	0.028	0.022	0.027	0.023
Var	0.015	0.002	0.003	0.011	0.011	0.005	0.004	0.008	0.014	0.032	0.009
Std	0.121	0.047	0.051	0.106	0.104	0.069	0.065	0.088	0.120	0.180	0.094
(c) Relative Size of Difference between Empirical and Revised Empirical LTSM											
Norm^2	0.004	0.001	0.000	0.002	0.002	0.001	0.000	0.001	0.000	0.001	0.001
Norm	0.062	0.028	0.017	0.042	0.041	0.029	0.018	0.026	0.022	0.028	0.029
Var	0.023	0.005	0.002	0.014	0.018	0.014	0.007	0.002	0.020	0.033	0.013
Std	0.151	0.074	0.045	0.118	0.135	0.118	0.085	0.048	0.142	0.181	0.112
(d) Relative Size of Difference between Empirical and Revised Empirical ATSM											
Norm^2	0.003	0.000	0.000	0.002	0.001	0.000	0.000	0.001	0.001	0.001	0.001
Norm	0.051	0.019	0.019	0.039	0.033	0.017	0.014	0.028	0.024	0.028	0.023
Var	0.015	0.002	0.003	0.012	0.012	0.005	0.004	0.010	0.016	0.035	0.010
Std	0.124	0.049	0.051	0.111	0.109	0.070	0.065	0.101	0.125	0.187	0.098

2.6 MODELS CONSISTENT WITH BLACK-SCHOLES THEORY

The covariance matrices obtained above can hardly be used to define a good state space Ω in which the Black-Scholes equation admits a unique solution and the sample path $\{(F_t^1, F_t^2) \mid t \in \mathbf{T}\}$ contains in Ω. According to Theorem 5, here we would like to modify the coefficients $\{\sigma_k^{ij}\}$ to allow us to have models with good state spaces.

2.6.1 The Revised Empirical LTSM Model

We choose

$$
\begin{aligned}
\sigma(x,y) &= \sigma_1 x + \sigma_2 y \\
&= \begin{pmatrix} 0.024 & 0 \\ 0 & 0.192 \end{pmatrix} x + \begin{pmatrix} 0 & 0.024 \\ 0.024 & -0.048 \end{pmatrix} y \\
&= \frac{1}{250} \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 4x+y & 0 \\ 0 & 2x-y \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}.
\end{aligned}
$$

These numerical values are in a vicinity of those empirical ones in (2.5.6) that we obtained from the sample path $\{(F_t^1, F_t^2)\}_{t\in\mathbf{T}}$. Using the above defined matrices σ_1 and σ_2 we solve the Riccati equations, obtaining the optimal parameters for the solution to match the empirical ones from SPCM. The result translates to the determination of the following functions:

$$
\begin{pmatrix} R(x,y) \\ P^1(x,y) \\ P^2(x,y) \end{pmatrix} = \begin{pmatrix} r_1 x + r_2 y \\ p_1^1 x + p_2^1 y \\ p_1^2 x + p_2^2 y \end{pmatrix} = \begin{pmatrix} 3.9821\, x - 1.0223\, y \\ 0.0788\, x + 0.0763\, y \\ -0.3056\, x - 0.4225\, y \end{pmatrix}. \tag{2.6.1}
$$

Here R is in the unit of %/year. Note that these values are almost identical to that in (2.5.9).

To see how well is the fit of the theoretical loads $L_k(P,s)$ from the solution of the

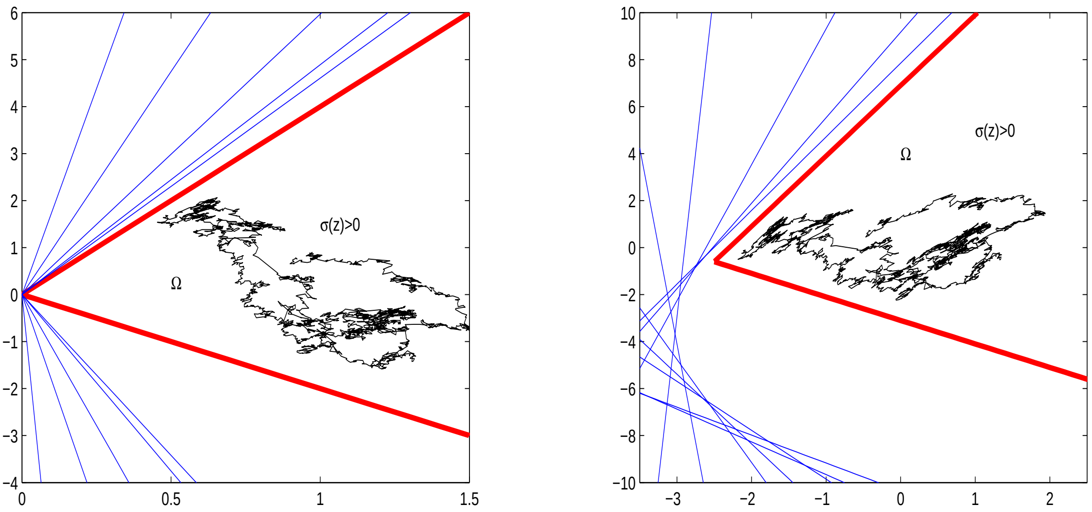
Riccati equations (2.2.6) and those $\ell_k(s)$ from 2-SPCM, we find that

$$\begin{aligned}\max_{k=1,2,s\in\mathbf{S}} |L_k(P,s)/s - \ell_k(s)| &= 0.14,\\ \left(\frac{1}{2m}\sum_{k=1}^{2}\sum_{s\in\mathbf{S}} |L_k(P,s)/s - \ell_k(s)|^2\right)^{1/2} &= 0.06.\end{aligned}$$

This produces an almost identical fantastic fitness as that depicted in Figure 7. The relative size of the difference between the empirical yield and the yield produce by this revised LTSM is listed in Table 3 (c). One finds that this table is almost identical to Table 3 (a).

Figure 9: The State Space Ω

(a) LTSM (b) ATSM



Thick half-lines are the boundary of Ω in which $\sigma(z) > 0$ and for each $s \in (0, 30]$, $(L_1(s), L_2(s)) \cdot z > 0$ for LTSM and $A_0 + (A_1(s), A_2(s)) \cdot z > 0$ for ATSM. Each thin line is given by the equation $(L_1(s^i), L_2(s^i)) \cdot z = 0$ for LTSM and $A_0(s) + (A_1(s), A_2(s)) \cdot z = 0$ for ATSM, $i = 1, \cdots, 10$. Also $\sigma(z)\mathbf{n}(z) = 0$ for $z \in \partial\Omega$. The Brownian motion like trajectory is the sample path $\{(F_t^1, F_t^2) \mid t \in \mathbf{T}\}$ for LTSM and $\{(X_t^1, X_t^2) \mid t \in \mathbf{T}\}$ for ATSM, which stays in Ω.

According the conditions needed in Theorem 5 and our special form of $\sigma(x, y)$, we now define

$$\Omega = \{(x, y) \in \mathbb{R}^2 \mid |x + y| < 3x\} \ .$$

It is easy to see the following:

$$\partial\Omega := \{(x, y) \mid x \geqslant 0, y = 4x\} \ \cap \{(x, y) \mid x \geqslant 0, y = -2x\},$$
$$\sigma(z) > 0 \quad \text{in} \quad \Omega, \qquad\qquad \sigma(z)\mathbf{n}(z) = \mathbf{0} \quad \forall\, z \in \partial\Omega.$$

Also, bond price $Z_t^{t+s} = \exp(-s[\ell_1(s)F_t^1 + \ell_2(s)F_t^2]) < 1$ is equivalent to $x > -\ell_2(s)/\ell_1(s)y$ for all $(x, y) \in \Omega$. Reading from (2-SPCM) in §2.5.5 and noting that the solution of the Riccati equations is very close to the empirical one, we see that

$$Z_t^{t+s} = \exp\Big(- L_1(s)F_t^1 - L_2(s)F_t^2\Big) < 1 \qquad \forall\, s \in (0, 30], \quad (F_t^1, F_t^2) \in \Omega.$$

The domain Ω and the empirical sample path $\{(F_t^1, F_t^2) \mid t \in \mathbf{T}\}$ are shown in Figure 9 (a).

2.6.2 The Revised Empirical ATSM Model

We choose

$$\begin{aligned}
\sigma(x, y) &= \sigma_0 + \sigma_1 x + \sigma_2 y \\
&= \begin{pmatrix} 0.30 & 0.07 \\ 0.07 & 1.04 \end{pmatrix} + \begin{pmatrix} 0.12 & 0 \\ 0 & 0.36 \end{pmatrix} x + \begin{pmatrix} 0 & 0.12 \\ 0.12 & 0.24 \end{pmatrix} y \\
&= \frac{3}{100} \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 3x - y + 69/10 & 0 \\ 0 & x + y + 31/10 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 3 \end{pmatrix}.
\end{aligned}$$

These numerical values are only in magnitude close to those empirical ones in (2.5.6) that we obtained from the sample path $\{(X_t^1, X_t^2)\}_{t\in\mathbf{T}}$. Using these matrices σ_0, σ_1 and σ_2 we solve the Riccati equations, obtaining the optimal parameters for the solution to match the empirical ones from PCM. The result translates to the determination of the following functions:

$$\begin{pmatrix} R(x,y) \\ \\ P^1(x,y) \\ \\ P^2(x,y) \end{pmatrix} = \begin{pmatrix} r_0 + r_1 x + r_2 y \\ \\ p_0^1 + p_1^1 x + p_2^1 y \\ \\ p_0^2 + p_1^2 x + p_2^2 y \end{pmatrix} = \begin{pmatrix} 0.0395 + 0.0175\,x - 0.0055\,y \\ \\ 0.3479 - 0.1592\,x + 0.2518\,y \\ \\ -0.2035 + 0.1397\,x - 0.2341\,y \end{pmatrix} \tag{2.6.2}$$

Note that these values are almost identical to that in (2.5.10).

To see the goodness of fit between the theoretical $A_k(P,s)$ of the solution of the Riccati 's equations (2.2.4) and those $a_k(s)$ from (2-PCM), we find that

$$\begin{aligned} \max_{k=1,2,s\in\mathbf{S}} |A_k(P,s)/s - a_k(s)| &= 0.13, \\ \left(\tfrac{1}{2m}\textstyle\sum_{k=1}^2 \sum_{s\in\mathbf{S}} |A_k(P,s)/s - a_k(s)|^2\right)^{1/2} &= 0.058. \end{aligned}$$

This produces an almost identical excellent fit as that depicted in Figure 7. The relative sizes of the difference between this revised empirical ATSM model and the empirical data are listed in Table 3(d) which is almost identical to Table 3 (b).

From our special choice of $\sigma(x,y)$, we now define

$$\Omega = \{(x,y) \in \mathbb{R}^2 \mid 3x > y - 69/10, x > -y - 31/10\}\ .$$

It is easy to see the following:

$$\partial\Omega := \{(x,y) \mid x \geqslant -5/2, y = 3x - 69/10\} \cap \{(x,y) \mid x \geqslant -5/2, y = -x - 31/10\},$$

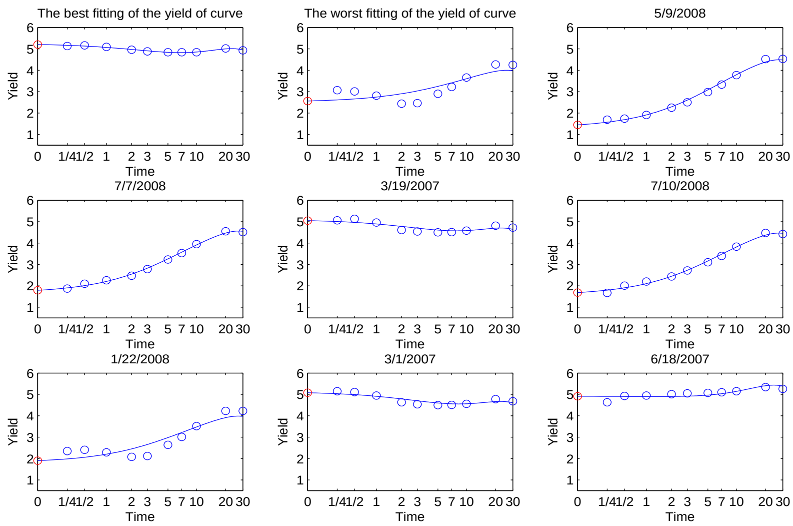
$$\sigma(z) > 0 \quad \text{in} \quad \Omega, \qquad\qquad \sigma(z)\mathbf{n}(z) = \mathbf{0} \quad \forall\, z \in \partial\Omega.$$

We can show that the $Z_t^T < 1$ for all $t \in (t, t+30]$ and that the empirical sample path $\{(F_t^1, F_t^2) \mid t \in \mathbf{T}\}$ is contained in Ω; see Figure 9 (b).

2.7 CONCLUSIONS AND EXTENSIONS

We obtained a two-factor Linear Term Structure Model and a two-factor Affine Term Structure Model that have the following properties: First, the state space Ω and parameters $(\sigma_k^{ij}, p_k^i, r_k)$ are chosen such that the Black-Scholes partial differential equation for pricing is completely determined and has a unique solution for every bounded payoff function. With the theoretical loads obtained from the solutions of the ordinary differential equations of Riccati type, the resulting bond price Z_t^T has the property that $Z_t^T < 1$ for $T \in (t, t+30]$. The empirical sample path stays in Ω. The yield surface produced from the LTSM and ATSM matches the empirical one at all points. Thus, we have two term structure models that are general, consistent, and accurate.

Figure 10: US Treasury Bond yield-to-Maturity curve 2007–2008



The ASTM loads are obtained by solving the Riccati equation.

The time period 1993–2007 is not very long. We would like to find data sets that cover longer period and expect the above conclusions continue to hold for models derived in the manner described in this paper. Now we let $\mathbf{T} = \{01/02/2007, ..., 08/29/2008\}$

be trading times and $\{Y_t^1\}_{t\in\mathbf{T}}, \cdots, \{Y_t^m\}_{t\in\mathbf{T}}$ be daily yield rate where m=10 as before. Thus, each column in the matrix $\{\mathbf{Y}_t\}_{t=1}^{419} = \{(Y_t^1, Y_t^2, ..., Y_t^{10})\}_{t=1}^{419}$ corresponds to the yield of 3 month, 6 month, 1 year, 2 year, 3 year, 5 year, 7 year, 10 year, 20 year, and 30 year bond, respectively.

The accuracy of the two–factor LTSM for the new dataset for 2007–2008 can be seen from Figure 10. In each plot in Figure 10, the dots are the actual yield and the curve is the fitted yield-to-maturity curve obtained by our LTSM. The first two plots are the best and worst fit respectively; the time t of all the 9 plots is randomly picked from our historical date set 2007–2008. Here the vertical axis has unit of percentage [3], and the horizontal axis is the index j of the maturity τ_j. One can see that for historical data, the two–factor LTSM, represented by the curve, fits the actual data, represented by the dots, very well.

A very interesting point to be noticed here is that we circled the red dots to represent the expected short–rates from the dynamic yield–to–maturity curves. The short rate is very important analytic basis for finance and macroeconomics. Duffie, et.al. and his followers used the assumption that the short–rates are affine in factors to build up the blocks of yields of bonds and other securities, which are risk–adjusted timely expected future short rates. And macro–economists emphasized the functions of short–rates in dictating the economic stabilities. However, the short–rate is itself random processes, which might be expected from some appropriate analysis. Here our LTSM gives a future testing direction for this expectations and extensions.

[3] Notice that the scale might be different for each plot, for the yield rates are changing over time.

3.0 LINEAR TERM STRUCTURE MODELS AND THE FORWARD PREMIUM ANOMALY

3.1 INTRODUCTION

The forward premium anomaly in the international currency markets refers to the fact that the forward rate is often found to be a biased estimator for the expected future spot rate in the future. This is puzzling because it implies that uncovered interest rate parity (UIP), a commonly assumed equilibrium condition does not hold in foreign exchange markets. Moreover researchers often find that when the forward rate is regressed on the future spot rate, the expected future rate under rational expectations, the coefficient is negative. When researchers study the forward premium anomaly, one possible explanation is the existence of a time varying risk premium.

Backus, Foresi, and Telmer (BFT) [13] (2001) examined the forward premium anomaly using affine term structure models. In their work, term structure models are adapted to a multi-currency setting. A term structure of interest rates is transferred to that of forward exchange rates through Covered Interest Parity (CIP). They use the affine models of Duffie and Kan [42] (1996), with the assumption for underlying random process following some specific random motion, then check the consistency of affine models with the anomaly equation posted by Fama [45] (1984), who argued that an explanation of the anomaly might be the presence of risk premium. They state that the Affine Term Structure Model (ATSM) [42] have difficulty in accounting for the anomaly. In their ATSM framework, they show that for the anomaly to exist at least one of the following two situations must hold: either interest rates are negative with positive probability, or the state variables have asymmetric effects on state prices of different currencies. Either alternative has important quantitative drawbacks. BFT formulated the prices as discrete random processes with different currency pricing kernels in their three specific cases and translated Fama's conditions for risk premium into restrictions on the affine parameters. Under the no-arbitrage condition, they use pricing kernels to derive the anomaly conditions based on the factors, or the underlying random processes. The underlying random processes are the basis for checking the affine parameters. However it is difficult to get hard evidence to test those restrictions in the models. The risk-neutral measure is used as a main assumption to show the proposition and derive the equations. On the other hand, factors can only be observed under the physical probability measure, whereas the risk-neutral probability measure is unobservable.

There are several contributions for this paper. First, I use the physical probability measure instead of the artificial risk-neutral measure (pricing kernels or discount factors), to derive a system of equations for the international currency interest rates and exchange rates. This is quite new and different compared to the previous literature. I model the behavior of the risk premium theoretically and empirically under the framework of the Linear Term Structure Model (LTSM) to study the forward premium anomaly. I also test my model using data on the Canadian-U.S. exchange rate. The dynamic factors are captured by Composite Principal Component Analysis (CPCA) which supplies a different way to set up the global factors for both currencies. The empirical results shows that two global factors can explain both American and Canadian interest rate quite well. Based on these factors, I can derive the parameters which are needed to solve the equations. The theoretical loads of LTSM are found by solving the Riccati ordinary differential equations, with the parameters chosen to match the ones from CPCA. It shows that the resulting LTSM presents interest-rate surfaces almost identical to the actual ones for both currencies, without imposing the symmetric restriction of factors on yields. Unlike previous work in this area, the theoretical interest rates are guaranteed to be positive.

Second, I put the theoretical results into the anomaly equation to test the effectiveness of LTSM. I find that the theoretical results can account for and reproduce the anomaly and the anomaly coefficients match the empirical coefficients quite well. Based on this, the conclusion can be drawn that LTSM can account for the Fama empirical findings. Furthermore, I derive the equations of the forward risk premium which are different from those of the existing literature. The extra risk part is firstly examined in this paper, whereas it has been vague in previous literature. The risk premium can be defined as the factors multiplied by the loading differences between domestic and foreign currencies, adjusted by the extra risk term. In this paper, the expected excess returns are represented by the risk premium associate with the risk-adjusted UIP relationship.

Finally, the risk premium captures part of the negative variance in the forward premium anomaly equation posted by Fama [45] (1984). As many researchers claimed that to account for the actual data, it requires theory to explain large fluctuations in risk premia, larger than those in the interest rate differentials. In this paper, I explain the larger fluctuations in risk premia by this extra risk term. My theoretical results are clearly consistent with the Fama's conditions: (1) negative covariance between risk premium and expected depreciation rate, and (2) greater variance of the first than the latter.

The rest of the paper is organized as follows. Section 2 reviews the theoretical models of exchange rates and international interest rates, and posts the questions which I will discuss in this paper. Section 3 provides a detailed derivation for the linear term structure of interest rates. I specify the global factors based on CPCA, and rebuild the exchange rate dynamics. Furthermore, I describes theoretical derivation for risk premia, and forward premia that shall be used. Section 4 presents the empirical factors and loads needed in LTSM; then I solve the Riccati equations to obtain theoretical loads for LTSM, with good state spaces. In Section 4, I use the empirical LTSM to examine the prediction power for exchange rate movement. Section 5 concludes the paper.

3.2 LITERATURE REVIEW

Some researchers illustrated that the potential shortcoming of the current international macroeconomic exchange rate models is that they may have incorrectly modeled the market's expectations on future changes of macroeconomic fundamentals. Since the yield structures have absorbed and shown the markets macroeconomic fundamental factors from different countries, we hold the point of view that the information contained in both the domestic and foreign term structures of interest rates are expected be simultaneously useful in accounting for exchange rate movements.

Duffie and Kan [42] (1996) systematically studied a special class of term structures, Affine Term Structure models(ATSM). The coefficients of the factors, called loads, are solutions of ordinary differential equations of Riccati type. They also suggested a discrete-time version to examine the pricing equation of ATSM, and posted the empirical results based on simulated GMM procedures. While Steenkiste and Foresi [87] (1999) imposed the Green's functions associated with the derivatives pricing for affine jump–diffusion process of Duffie and Kan [42] (1996) and illustrated their numerical implementations, Dai, Le, and Singleton [34] (2006) develops a rich class of discrete-time, nonlinear dynamic term structure models (DTSMs), nesting the exact discrete-time counter-parts of Duffie and Kan [42] (1996) and Dai and Singleton (2000), and emphasizing the physical probability measure.

The pioneering work in this area are by Fama [45] (1984), Amin and Jarrow (1991), Backus et al. (1995), Ahn (1997), Bakshi and Chen (1997), Basal (1997). Fama [45] (1984) examined whether there exists a time-varying risk premium and tested whether forward rates contain information on future spot rates. Given market efficiency or rationality, Fama's decomposition let the forward rate be interpreted as the sum of a premium and an expected future spot rate. He concluded that the reason why the slope coefficients in the expected future spot rate regressions is negative is that the variance of the premium of forward rate is much larger than the variance of expected future spot rate. The premium in the forward rate expression equation is just the difference between the expected real returns on the nominal interest rates of the two countries. Thus, the factors that determine the difference of interest rates will also determine the premium in the forward rates, and, furthermore, will explain variation in forward premium and expected future spot rate. My research exactly followed this idea. There is a promising standpoint to develop empirically applied, internally consistent models of cross–country term structures on exchange rates. Backus, Foresi, and Telmer [13] (2001) examined the forward premium anomaly in the affine term structure models. The term structure models imply a term structure of forward exchange rates in the anomaly equation as long as the models are adapted to a multi-currency setting. They state that the Affine Term Structure Models (ATSM) [42] have difficulties accounting for the forward premium anomaly, either allowing for the negative theoretical interest rate with positive probability, or for asymmetric effects of state variables on interest rates for different countries. They examined the forward premium anomaly in the context of affine models of the term structure of interest rates, and find the quantitative properties of either alternative have important drawbacks. They also suggested a discrete-time version to use the pricing equation of ATSM, and posted the empirical results based on simulated GMM procedures.

I open my questions by revisiting Backus, Foresi, and Telmer [13] (2001). The empirical regression equations are of the form

$$s_{t+1} - s_t = a_1 + a_2(f_t - s_t) + residual, \tag{3.2.1}$$

where s is the logarithm of spot exchange rate, f is the logarithm of forward exchange rate.

They followed Fama (1984) in decomposing the forward premium

$$f_t - s_t \quad = \quad (f_t - \mathbb{E}_t s_{t+1}) + (\mathbb{E}_t s_{t+1} - s_t) = p_t + q_t, \tag{3.2.2}$$

in which q_t is the expected rate of depreciation, and p_t is the risk premium as interpreted by many researchers. The population regression coefficients

$$a_2 \quad = \quad \frac{\text{cov}(q, p+q)}{\text{var}(p+q)} = \frac{\text{cov}(q,p) + \text{var}(q)}{\text{var}(p+q)}, \tag{3.2.3}$$

Backus et al. [13] (2001) derive the empirical setup for the short-rate as

$$r_t = -\log \mathbb{E}_t^* m_{t+1} = -\big(\mathbb{E}_t \log m_{t+1} + \frac{1}{2}\text{var} \log m_{t+1}\big). \tag{3.2.4}$$

where m is the dollar pricing kernel. Here, I put $*$ on $\mathbb{E}_t^*$ to let reader notice that this expectation is under no-arbitrage condition. $\mathbb{E}_t$ is the expectation under the physical probability measure, and surely var_t is the second moment under the physical probability measure too. They derive the depreciation rate under no-arbitrage condition, further examine the anomaly equation, and then make derivation for affine parameters. Under no arbitrage assumption, risk-neutral probability and pricing kernels are used in their work. Furthermore, $\mathbb{E}_t \log m_{t+1}$ and $\frac{1}{2}\text{var} \log m_{t+1}$ are calculated based on their assumption of a specific processes for the state variables, factors or the underlying driving random elements. This assumption may be redundant, because there is no need to restrict the state variables themselves to follow a specific distribution, as long as the factors can be derived from the observable data and the pricing model works well with them. Many researchers follow their derivation in empirical testing for ATSM.

To extend more based on their works, I had been thinking of the following several questions. Backus et al. [13] (2001) suggested three restriction conditions (case A, B and C) for factors in their specific two-currency model: independent factors, one common factor and independent factors, and interdependent factors. For their case

A, there are two questions here: whether it is necessarily to restrict the coefficients associated with the factors from one currency are equal to those from the other currency. And whether it is necessarily true that two countries pricing kernels are independent. For Case B, what roles do the common factors play on have the effects on both currencies. The empirical results are needed to testify it. For Case C, they think the last model may be better than the other two models, but there exist the difficulties that the unconditional distribution of the interest rates do not exhibit the extreme behavior of the factors. In this paper, I try to find some new ways to answer the above questions.

There have been many other researchers developing lots of exciting works in this area. While Steenkiste and Foresi [87] (1999) imposed the Green's functions associated with the derivatives pricing for affine jump–diffusion process of Duffie and Kan [42] (1996) and illustrated their numerical implementations, Dai, Le, and Singleton [34] (2006) develops a rich class of discrete-time, nonlinear dynamic term structure models (DTSMs), nesting the exact discrete-time counter-parts of Duffie and Kan [42] (1996) and Dai and Singleton (2000), and emphasizing the physical probability measure. However, how to directly deal with the physical probability measure for Affine Term Structure models, is still not set up for the empirical applications, especially in explaining the forward premium anomaly. Furthermore, there are many followers of BFT(2001). Ahn [3] (2004) develops two-country term structure and exchange rate pricing and examine a diversification effect for an international bond portfolio, although the physical probability measure is not explicitly discussed. The common and local factors are set up by orthogonal condition based on the preliminary results from PCA. Inci and Lu [64] (2004) follow the quadratic class of Ahn et. al. [2] (2001), and extended Backus et al. (2001) to allow for a more flexible conditional correlations structure among state variables in the simulated joint factors, and nominal interest rates are guaranteed to be positive. They pointed out the empirical performance in tracking movements of exchange rates and currency returns term structure model.

The forward premium puzzle is explained, but exchange rates are also affected by other factors that are not in the interest rate dynamics. Junker, Szimayer, and Wagner [66] (2006), based on a two–factor generalized Vasicek ATSM model [42, 11] and copula, developed a nonlinear cross-sectional dependence among US treasury bonds. While theoretically well–studied, the ATSM was empirically studied by Piazzesi [81] (2003), with the help of Principal component analysis (PCA). All of these followers made their discussion under the framework of BFT(2001).

To set up term structure models in multi-country environment, there are some other things need to be concerned: how to model the stochastic processes; what kind of factors or state variables are selected; and whether some assumption is necessary. Backus, Foresi and Telmer [13] (2001) assumed three cases based on different properties of factors which were used in their empirical testing. They used Cox, Ingersoll, and Ross (CIR)(1985) model as one example of ATSM under the risk–neutral measure to account for the properties of currency prices and interest rates in three different cases. They showed that the models with interdependent factors, or global factors seem to render the most striking results for the ATSM capturing the properties of currency prices and their interest rates, but the results are not implemented using the actual data and have big shortcomings mentioned in their conclusion. Dewachter and Maes [36] (2001) estimate continuous-time multi-factor ATSM for the interest rate dynamics across countries, and incorporate the exchange rate dynamics to examine the forward premium puzzle. The local factors are derived using nonlinear optimization, combined with the preliminary PCA results based on two-country datasets. The transformation produces deviation inside the variance matrix, and the derived common factor needs more discussion. They also alternatively supplied the factors on the stimulated joint distributions with Kalman filter algorithm. The results are drawn based on BFT(2001), which could be checked again. Ahn [3] (2004) derived the common and local factors by orthogonal condition based on the preliminary results from PCA, similar to the first method above, and Inci and Lu [64] (2004) make a quite similar simulation as the latter method above. Perignon, et. al. [79] (2007) analyzed country-specific factors when estimating common factor structure of US, German, and Japanese Government bond returns. They use so–called inter–battery factor analysis model suggested by Tucher (1958) in psychology to show that the classical PCA on a multi-country dataset of bond returns vaguely captures both local and common factors, and conclude that US bond returns share only one common factor with German and Japanese bond returns. All of these empirical works were based on BFT(2001), although they illustrate many different steps for setting their PCA models.

3.3 MODEL FRAMEWORK AND EXTENSION

3.3.1 Exchange Rate Dynamics

The flexible-price (Monetarist) monetary model developed by Frenkel (1976,1977,1980) assumes that all goods prices are flexible both in long-run and short-run, that capital is perfectly mobile, and that domestic and foreign assets are perfect substitutes, which is under the theory framework of the asset-market view of exchange rate. So the exchange rate must adjust instantly to equilibrate the international demand for stocks of national assets, as long as the purchasing power parity (PPP) holds continuously. This empirical implication is that floating exchange rates will exhibit high variability, which exceeds what one might regard as that of their underlying determinants. Under the views of asset-market, when assumed that domestic and foreign bonds are perfect substitutes: Portfolio shares are infinitely sensitive to expected rates of the return, thus uncovered interest parity (UIP) holds. Given that it does hold, bond supplies then become irrelevant. The responsibility for determining the exchange rate is shifted onto the money markets. Such models belong to the monetary approach to exchange rate, focusing on the demand for and supply of money, and other important fundamentals. Because of asymmetric information transmission, the market is always not perfect market. The risk premium exists because the assets are imperfect substitutes. Frenkel (1993) said that the speculative bubble of the exchange rate can be involved into the equation with the fundamentals as the approximation of risk premium.

The results of empirical models are disappointing, reported in Frankel (1993), MacDonald and Taylor (1992). The monetary model estimation procedures suffer from some serious deficiencies. On the theoretical side, in the conventional monetary model, the exchange rate adjusts to balance the international demand and supply of monetary assets. The demand for money is always considered to be a function of the level of interest rates and income. Backus, et. al. [13] (2001) examine the risk premium under the asset pricing view. The model framework and extension will be given in this section by LTSM.

3.3.2 Uncovered Interest Parity and Covered Interest Parity

If the investors can cover the investment with a forward contract, the arbitrage between two investment opportunities results in a CIP condition. Let F_t^T, S_t be the forward and spot exchange rates (units of US currency per unit of foreign currency) observed at time t, and let R_t^T and R_t^{T*} be the nominal interest rates observed at time t on eurocurrency interest rates or discount bonds denominated in US currency and foreign currency. The bonds are zero-coupon bonds with the same maturity for each country.

$$1 + R_t^T = \frac{F_t^T}{S_t}\left(1 + R_t^{T*}\right). \tag{3.3.1}$$

so the interest rate parity equation can be approximated by

$$r_t^T - r_t^{T*} = f_t^T - s_t. \tag{3.3.2}$$

here $r_t^T, r_t^{T*}, f_t^T, s_t$ represent the logarithm of $1 + R_t^T, 1 + R_t^{T*}, F_t^T, S_t$, which says that interest differential between a US denominated investment instrument and a Canadian denominated investment is equal to the forward premium or discount on the Canadian. On the other hand, UIP conditions says that the expected level of spot rate is equal to forward rate, so the expected change in spot exchange rate is equal to the forward premium or discount, so to the interest differential. Extensive studies have shown that UIP does not hold in the cross-country data. Furthermore, the forward premium anomaly states that the expected spread in spot exchange rate is negatively related to the forward premium, so, in other words, higher interest rate currencies are more likely to appreciate.

3.3.3 Linear Term Structure Model

The model setup begins with linear functions in the sense of combinations of factors which are any stochastic processes. A term structure models time t value of T-bond for any $t \geqslant 0$ and $T > t$. Here we use B_t^T to denote the price of one share of T-bond at time t. By default, $B_T^T = 1$. Following Chen and Huang [24] (2008), we firstly assume

(A1) *the collection* $\{B_t^T\}_{0 \leqslant t \leqslant T, T>0}$ *obeys a stochastic differential equation*

$$\frac{dB_t^T}{B_t^T} = \mu_t^T \ dt + \sum_{i=1}^{n} \sigma_{t\,i}^T \ \ dX_t^i \tag{3.3.3}$$

where $dB_t^T = B_{t+dt}^T - B_t^T$, $\{(X_t^1, \cdots, X_t^n)\}$ *is a stochastic process and* $\{\mu_t^T, \sigma_{t1}^T, \cdots, \sigma_{tn}^T\}$ *are stochastic processes adapted to a natural filtration*[4].

Assume **(A1)** in an arbitrage-free system. According to Chen and Huang [24] (2008),then there exist processes $\{R_t, P_t^1, \cdots, P_t^n\}$ adapted to a natural filtration

[4] A process $\{x_t\}_{t \in \mathbf{T}}$ is **adapted to a natural filtration** if x_τ is observable at any time $t \geqslant \tau \in \mathbf{T}$.

such that

$$\mu_t^T = R_t + \sum_{i=1}^{n} P_t^i \, \sigma_{t\,i}^T \qquad \forall\, T > 0, t \in [0, T]. \tag{3.3.4}$$

Here short-term bonds can be treated as risk-free in the sense that buying a $(t+dt)$-bond at time t and selling it at time $t+dt$ produces a fixed return rate, R_t, called the **short-rate**. This assumption can be implemented (e.g. [57]) by assuming $\sigma_{t\,i}^{t+dt} = 0$ for all i, t, therefore (3.3.8) gives $R_t = \mu_t^{t+dt}$ and (3.3.3) implies $B_t^{t+dt} = e^{-R_t dt}$. Typically $\{(X_t^1, \cdots, X_t^n)\}$ in (3.3.3) is assumed to be a martingale under a measure $\mathbb{P}$ of physical observations, hence μ_t^T is the observed expected return rate of the investment on T-bond. The identity (3.3.8) further claims that any increment of the expected return beyond the short-rate R_t can only be achieved with risks. In (3.3.8), the P_t^i of the volatility $\sigma_{t\,i}^T$ is therefore called the **price of risk** on the **uncertainty innovation** $\sigma_{ti}^T dX_t^i$.

The linear term structure model, **LTSM**, assumes that the logarithms of bond prices are linear functions of factors; i.e,

(A2) *The price B_t^T of the T-bond at time t satisfies*

$$\log \frac{1}{B_t^T} = \sum_{i=1}^{n} L_i(T-t)\; F_t^i \quad \forall\, t \geqslant 0, T \in [t, t + T^{\max}), \tag{3.3.5}$$

where $L_1(\cdot), \cdots, L_n(\cdot)$ are differentiable functions defined on $[0, T^{\max})$ and $\{(F_t^1, \cdots, F^n)\}$ is an Itô process with a positive definite matrix $(\mathbb{E}[F_\tau^i F_\tau^j])_{n\times n}$ for some $\tau \geqslant 0$. Along with

(A3) *We have an arbitrage–free environment.*

there are constants $c_k, p_k^i, \sigma_k^{ij} = \sigma_k^{ji}$ for $k, i, j = 1, \cdots, n$ such that the functions $L_1(\cdot), \cdots, L_n(\cdot)$ are solutions of the **Riccati** equations

$$\begin{cases} \dfrac{dL_k(s)}{ds} &= c_k - \displaystyle\sum_{i=1}^{n} p_k^i L_i(s) - \frac{1}{2} \sum_{i,j=1}^{n} \sigma_k^{ij} L_i(s) L_j(s) \qquad \forall\, s \in [0, T^{\max}), \\ L_k(0) &= 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k = 1, \cdots, n. \end{cases} \tag{3.3.6}$$

If $L_i(\cdot), L_i(\cdot)L_j(\cdot), i = 1, \cdots, n, j = 1, \cdots, i$, are linearly independent, then

$$\frac{dB_t^T}{B_t^T} = R_t\, dt - \sum_{i=1}^{n} L_i(T-t) \left\{ P_t^i dt + dF_t^i \right\}, \tag{3.3.7}$$

where process $\{R_t, P_t^1, \cdots, P_t^n\}$ adapted to a *natural* filtration such that

$$\mu_t^T = R_t + \sum_{i=1}^{n} P_t^i\, \sigma_{t\,i}^T \qquad \forall\, T > 0, t \in [0, T],$$

$$R_t = \sum_{k=1}^{n} c_k F_t^k, \qquad P_t^i = \sum_{k=1}^{n} p_k^i F_t^k, \qquad \frac{\mathrm{Cov}(dF_t^i, dF_t^j)}{dt} = \sum_{k=1}^{n} \sigma_k^{ij} F_t^k\ . \tag{3.3.8}$$

Notice that **LTSM** generalizes **ATSM** by allowing time dependent mean returns. We can write (3.3.5) as

$$\log \frac{1}{B_t^T} = L_0(T-t, t) + \sum_{k=1}^{m} A_i(T-t)\, X_t^i$$

$$m_t^i = \mathbb{E}[F_t^i], \qquad X_t^i = F_t^i - m_t^i, \quad L_0(s,t) := \sum_{i=1}^{n} L_i(s) m_t^i, \quad A_i(s) = L_i(s),$$

When $\{(F_t^1, \cdots, F_t^m)\}_{t\in\mathbb{R}}$ is stationary, m_t^k does not depend on t, so $L_0(s,t)$ depends only on s and LTSM is a special ATSM in which $A_0(\cdot)$ is a linear combination of $A_1(\cdot), \cdots, A_n(\cdot)$.

3.3.4 Important Differences for the Model Setup

The $\{(X_t^1, \cdots, X_t^n)\}$ is a stochastic process, which is observable. μ_t^T is the observed expected return rate of the T-bond, and any increment of the expected return from the short-rate R_t can only be achieved with risks. We can model any T-bond, not just the short-term rate. The pricing kernel for short-rate, or risk-free rate, is not used in this paper, here, B_t^{t+1} may be equivalent to the $\mathbb{E}_t^* m_t^{t+1}$, with $dt = 1$ in discrete version, since $\int_0^t P_s^i ds$ is zero, X_t^{*i} can be treated as X_t^i, so we don't have to bother the increments from the different measures.

3.3.5 Forward Premium Examination

This section incorporates the exchange rate dynamics in the framework of asset–pricing view to completely pin down the theoretical environment for examining the forward premium puzzle. Now we consider the SDE or state equation or transition function [5] :

$$\frac{dB_t^T}{B_t^T} = R_t\, dt - \sum_{i=1}^{n} L_i(T-t)\left\{P_t^i dt + dF_t^i\right\}, \tag{3.3.9}$$

and, accordingly,

$$\frac{dB_t^{T*}}{B_t^{T*}} = R_t^*\, dt - \sum_{i=1}^{n} L_i^*(T-t)\left\{P_t^{i*} dt + dF_t^i\right\}. \tag{3.3.10}$$

I construct the determination of exchange rates and currency returns in the context of the linear term structure model (LTSM) based on the above state equation or transition function.

According to Chen and Huang [24] (2008), there are constants $c_k, p_k^i, c_k^*, p_k^{i^*}, \sigma_k^{ij} = \sigma_k^{ji}$ for $k, i, j = 1, \cdots, n$ such that the functions $L_1(\cdot), \cdots, L_n(\cdot)$, and $L_i^*(\cdot), \cdots, L_n^*(\cdot)$ are solutions of the **Riccati** equations

$$\begin{cases} \dfrac{dL_k(s)}{ds} & = c_k - \displaystyle\sum_{i=1}^{n} p_k^i L_i(s) - \frac{1}{2}\sum_{i,j=1}^{n} \sigma_k^{ij} L_i(s) L_j(s) \qquad \forall\, s \in [0, T^{\max}), \\ L_k(0) & = 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad k = 1, \cdots, n. \end{cases} \tag{3.3.11}$$

[5] where if associated with short-rate expectation, R_{t+1} as the gross one-period short-rate return, m as the pricing kernel, $\mathbb{Q}$ as the risk-neutral measure, based on the information at time t and assuming no-arbitrage condition:

$$1 = \mathbb{E}_t^{\mathbb{Q}}(m_{t+1}R_{t+1})$$

similarly the pricing kernel for foreign currency:

$$1 = \mathbb{E}_t^{\mathbb{Q}}(m_{t+1}^* R_{t+1}^*) = \mathbb{E}_t^{\mathbb{Q}}\big(m_{t+1}(S_{t+1}/S_t)R_{t+1}^*\big)$$

but it is under risk–neutral measure, and can be translated to our interests under physical probability measure.

and

$$\begin{cases} \dfrac{dL_k^*(s)}{ds} & = c_k^* - \displaystyle\sum_{i=1}^{n} p_k^{i^*} L_i^*(s) - \frac{1}{2}\sum_{i,j=1}^{n} \sigma_k^{ij} L_i^*(s) L_j^*(s) \qquad \forall\, s \in [0, T^{\max}), \\ L_k^*(0) & = 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad k = 1, \cdots, n. \end{cases} \tag{3.3.12}$$

If $L_i(\cdot), L_i(\cdot)L_j(\cdot), i = 1, \cdots, n, j = 1, \cdots, i$, are linearly independent, and $L_i^*(\cdot), L_i^*(\cdot)L_j^*(\cdot), i = 1, \cdots, n, j = 1, \cdots, i$, are also linearly independent, for both currencies, then

$$\frac{dB_t^T}{B_t^T} - \frac{dB_t^{T*}}{B_t^{T*}} = (R_t - R_t^*)\, dt - \sum_{i=1}^{n} L_i(T-t) \left\{ P_t^i dt + dF_t^i \right\} + \sum_{i=1}^{n} L_i^*(T-t) \left\{ P_t^{i*} dt + dF_t^i \right\},$$

where process $\{R_t, P_t^1, \cdots, P_t^n\}$ adapted to a *natural* filtration such that

$$\mu_t^T = R_t - R_t^* + \sum_{i=1}^{n} P_t^i \, \sigma_{t\,i}^T - \sum_{i=1}^{n} P_t^{i*} \, \sigma_{t\,i}^T \quad \forall\, T > 0, t \in [0, T],$$
$$R_t = \sum_{k=1}^{n} c_k F_t^k, \qquad P_t^i = \sum_{k=1}^{n} p_k^i F_t^k,$$
$$R_t^* = \sum_{k=1}^{n} c_k^* F_t^k, \qquad P_t^{i*} = \sum_{k=1}^{n} p_k^{i*} F_t^k, \qquad \frac{\mathrm{Cov}(dF_t^i, dF_t^j)}{dt} = \sum_{k=1}^{n} \sigma_k^{ij} F_t^k\ . \tag{3.3.13}$$

Theoretically speaking, the R_t and R_t^* are of the same value if they are treated as the short rates, because, under no-arbitrage condition, the short rates are equivalent across the country to guarantee there is no arbitrage opportunity. To show that the resulting exchange rate dynamics incorporated with LTSM present exchange–rate surfaces almost identical to the empirical one for both currencies, and LTSM can explain the forward premium anomaly with suitable state space, I switch to discrete-time implementations and fix the sampling frequency so that the time interval is, some $\tau \geqslant 0$ for both currencies, say, one working day.

3.3.6 Expected Rate of Depreciation Examination and Risk Premium Setup

Under Covered Interest Parity (CIP) and the linear term structure model (LTSM), according to Appendix A.5, the price B_t^T of the T-bond at time t satisfies

$$f_t^T - s_t = r_t^T - r_t^{T*} = \frac{1}{T-t}\sum_{i=1}^{n} L_i(T-t)\; F_t^i - \frac{1}{T-t}\sum_{i=1}^{n} L_{i*}(T-t)\; F_t^i, \qquad (3.3.14)$$

When associated with spot-rate expectation and the rate of depreciation expectation:

$$s_t = fm_t + \lambda(\mathbb{E}s_{t+1} - s_t). \qquad (3.3.15)$$

where fm_t are the economic fundamentals, and this is the basis for setting up the model for expected currency assets. Using the state equations of (3.3.9) and (3.3.10), and the associated yield equations using the linear term structure model, **LTSM**, the price B_t^T of the T-exchange at time t satisfies

$$r_t^T = -\frac{1}{T-t}\log\frac{1}{B_t^T} = \frac{1}{T-t}\sum_{i=1}^{n} L_i(T-t)\; F_t^i \quad \forall t \geqslant 0, \qquad (3.3.16)$$

and, accordingly,

$$r_t^{T*} = -\frac{1}{T-t}\log\frac{1}{B_t^{T*}} = \frac{1}{T-t}\sum_{i=1}^{n} L_i^*(T-t)\; F_t^i \quad \forall t \geqslant 0. \qquad (3.3.17)$$

where $T \in [t, t+T^{\max})$, and $\{(F_t^1, \cdots, F^n)\}$ is an Itô process, and global factor for both currency assets.

Recall that $\{(F_t^1, \cdots, F_t^n)\}$ is assumed to be a martingale under a measure $\mathbb{P}$ of physical observation. In Appendix A.6, $S_T^{\mathbb{Q}}$ can be written as

$$S_T^{\mathbb{Q}} := S_T + \int_0^T K_\zeta d\zeta \qquad \forall T \geq t \geqslant 0. \qquad (3.3.18)$$

Suppose $\mathbb{Q}$ is a measure under which $\{\{S_T^{\mathbb{Q}}\}\}_{T\geq t}$ are martingales. This particular measure $\mathbb{Q}$ is called the *risk-neutral measure*. The K_ζ is firstly examined in this paper, named as K-risk. This is the extra risk part we need to examine under the physical probability measure. It is a useful risk part, which is vague in previous literature and can be used to explain the risk which cannot be explained in the past years. The future spot exchange rate, S_T, is unknown at time t, so it's risky for those investors who get dollar returns on the foreign assets. To compensate the investor who chooses to hold the foreign assets for taking on this exchange rate risk, the K-risk, K_ζ, is added into equation (A.6.3) to guarantee the investor, who is risk-averse, get the same return as that from dollar assets.

Following the derivation in Appendix A.6 and what we derived above, and recalling the decomposition equation (3.2.2), the expected rate of depreciation can be expressed as

$$\mathbb{E}\ln S_T - \ln S_t = \mathbb{E}(\ln S_T - \ln S_t)$$
$$\mathbb{E}\ln S_T - \ln S_t = \mathbb{E}\ln S_T - \mathbb{E}\ln(S_T + \int_t^T K_\zeta^i d\zeta) + \sum_{i=1}^{n}[L_i(T-t) - L_i^*(T-t)]\ \mathbb{E}F_t^i \tag{3.3.19}$$

we have expected rate of depreciation denoted by the mean path of the factors multiplied by the loading difference between domestic and foreign currencies, adjusted by the K-risk. So we have the risk premium as

$$f_t - \mathbb{E}\ln S_T = \sum_{i=1}^{n}[L_i(T-t) - L_i^*(T-t)]\ [F_t^i - \mathbb{E}F_t^i] + \mathbb{E}\ln(S_T + \int_t^T K_\zeta^i d\zeta) - \mathbb{E}\ln(S_T) \tag{3.3.20}$$

where $\mathbb{E}_t F_t^i$ can be calculated by $\int_0^t F_t^i dt$, and as before we still don't put any assumption on the probability distribution of F_t^i. The risk premium can be defined as randomness of the factors multiplied by the loading difference between domestic and foreign currencies, adjusted by the K-risk.

Here, there is more room for us to discuss again the form of risk premium we derived above. From the beginning, we set up the term structure model for the process of interest rates. Although we stick with the physical probability measure for empirical settings, we still show the extra terms under the risk-neutral measure.

Again, the risk premium we derived under the term structure also has an extra term, which is consistent with what we discuss at the beginning of the model setup.

About exchange rate changes, the excess return, defined here for the foreign currency exchange rate, is the difference in the cross-country yields adjusting for the relative currency movements, or the percent appreciation of foreign currency. As discussed earlier, in the efficient and complete foreign exchange market, if investors are assumed to be risk neutral and have rational expectations, expected exchange rate changes are equal to cross-country interest rate differences over the same horizon. This is the Uncovered Interest Parity (UIP) condition. However, the UIP condition is systematically not true over a wide rage of currency-interest rate pairs. In this paper, the expected excess returns in equation (3.3.20) represents the risk premium associate with K-risk in the risk-adjusted UIP relationship.

Consider the risk faced by an investor in USA who chooses between bonds denominated in either US dollars or Canadian dollars. His dollar return on the Canadian bond is risky because he does not know the next period's exchange rate at this moment. The risk premium compensates the investor who chooses to hold the Canadian bond for taking on this exchange rate risk. The risk premium captures part of the negative variance in the forward premium anomaly equation posted by Fama [45] (1984). As many researchers claimed that data requires theory to explain large fluctuations in risk premia, larger than those in the interest rate differentials. According to this section's discussion, equation (3.3.19) explain the larger fluctuations in risk premia, and this is also the adjustment for the relative currency movements, or the percent appreciation of foreign currency. To get a negative value of a_2, we need $cov(q,p) + var(q) < 0$, consistent with the Fama's conditions: (1) negative covariance between p and q, and (2) greater variance of p than q. In equation (3.3.20), $E_t F_t^i$ can be calculated by $\int_0^t F_t^i dt$. p and q are given by equation (3.3.20) and equation (3.3.19). Clearly, p and q are negatively correlated. To check condition (2), note that $\sum_{i=1}^n [L_i(T-t) - L_i^*(T-t)]\ F_t^i$ is the extra random term in equation (3.3.20), compared with equation (3.3.19), so the variance of p must be greater than q.

One of the central issue in term structure model is the characterization of the stochastic process $\{X_t^i\}$. We shall define them in terms of statistical common factors. Here, for both currencies, statistically it means factors are obtained from the given random variables themselves. I use Composite Principal Component Analysis (CPCA) to construct the empirical factors and loadings for further theoretical derivations.

3.3.7 Factors and Loads

In this section, similar to Chen and Huang [24] (2008),I use Composite Principal Component Analysis (CPCA) to construct the empirical factors and loadings for further theoretical derivations. This is also one of the central issue in term structure model to characterize the stochastic process $\{X_t^i\}$. From the given random variables themselves, I shall define the global factors and common factors, and their associated loadings or coefficients accordingly for both currencies.

3.3.7.1 Composite Principal Component Analysis Let $\left\{\{\xi_j^i\}_{j=1}^2\right\}_{i=1}^N$ be the American and Canadian bond random yields curve. Let fix i, the time-to-maturity periods as $1/4, 1/2, \cdots, 5$ years. i.e. ten different international bonds are considered here. I set $\left\{\{\xi_j^i\}_{j=1}^2\right\}_{i=1}^N$ as points in a Hilbert space $(\mathbf{H}, \langle\cdot,\cdot\rangle)$. Let $N = 1, \cdots, 6$, and $\left\{\{\lambda_j^i\}_{j=1}^2\right\}_{i=1}^N$, arranged in decreasing order with respect to index j, be a complete set of eigenvalues of matrices $\{\mathbf{C}_i\}_{i=1}^N$ where $\mathbf{C}_i := (\langle\xi_i^m, \xi_i^n\rangle)_{2\times 2}$. Let $K = \dim(\{\xi^1, \cdots, \xi^N\})$, then we construct $\left\{\{F_j^1, \cdots, F_j^K\}\right\}_{j=1}^2$, a set of principal components of $\{\xi_1^1, \cdots, \xi_1^N; \xi_2^1, \cdots, \xi_2^N;\}$. I denote the associated eigenpairs as $\{\langle\lambda_j^i, \mathbf{e}_j^i\rangle_{j=1}^2\}_{i=1}^N$, where $\{\lambda_1^i, \lambda_2^i\}$ is in a decreasing order. Thus, $\left\{\{F_j^1, \cdots, F_j^K\}\right\}_{j=1}^2$ is a set of composite principal components of $\{\xi_1^1, \cdots, \xi_1^N; \xi_2^1, \cdots, \xi_2^N;\}$. if and only if there exist row vectors $\mathbf{e}_1, \cdots, \mathbf{e}_K$ in $\mathbb{R}^N$, $\mathbf{e}_k = (e_k^1, \cdots, e_k^N)$, such that

$$\mathbf{e}_k \mathbf{C} = \lambda_k \, \mathbf{e}_k, \qquad \mathbf{e}_k \cdot \mathbf{e}_l = \delta_{kl}, \qquad F_j^k = \frac{1}{\sqrt{\lambda_k}} \sum_{k=1}^{2} \sum_{i=1}^{N} \xi_j^i \, \mathbf{e}_j^i(k) \quad \forall\, k, l = 1, \cdots, K.$$

Now only $\{\langle \lambda_1^i, \mathbf{e}_1^i \rangle\}_{i=1}^N$ are selected to be considered to construct global factors, $\{F_1^i\}_{i=1}^N$, which can be used to construct the composite Principal components analysis (CPCA) as the follows: Let the $\{\langle \eta_j^i, \mathbf{f}_j^i \rangle_{j=1}^2\}_{i=1}^N$ be the complete set of eigenpairs of matrix $\mathbf{C}_i := (\langle F_1^m, F_1^n \rangle)_{N \times N}$ with $\{\eta^i\}_{i=1}^N$ decreasing order. Thus, the composite global Principal component, $\{X^i\}_{i=1}^N$ as

$$X_j = \sum_{i=1}^{N} F_1^i f_j^i = \sum_{i=1}^{N} \left[\left(\sum_{k=1}^{2} \xi_1^i \, \mathbf{e}_1^i(k) \right) f_j^i \right] = \sum_{i=1}^{N} \sum_{k=1}^{2} \xi_1^i \, \mathbf{e}_1^i(k) f_j^i \qquad \forall\, k, l = 1, \cdots, K.$$

Proposition 1. *Let* $\left\{ \{\xi_j^i\}_{j=1}^2 \right\}_{i=1}^N$ *be points in a Hilbert space* $(\mathbf{H}, \langle \cdot, \cdot \rangle)$ *and* $\{f_i\}_{i=1}^N$*, arranged in decreasing order, be a complete set of eigenvalues of matrices* $\{\mathbf{C}_i\}_{i=1}^N$ *where* $\mathbf{C}_i := (\langle \xi_i^m, \xi_i^n \rangle)_{2 \times 2}$*. Let* $K = \dim(\{\xi^1, \cdots, \xi^N\})$*, then we construct* $\left\{ \{F_j^1, \cdots, F_j^K\} \right\}_{j=1}^2$*, a set of principal components of* $\{\xi_1^1, \cdots, \xi_1^N; \xi_2^1, \cdots, \xi_2^N; \}$*. Let denote the associated eigenpairs as* $\{\langle \lambda_j^i, \mathbf{e}_j^i \rangle_{j=1}^2\}_{i=1}^N$*, where* $\{\lambda_1^i, \lambda_2^i\}$ *is in a decreasing order. Thus,* $\left\{ \{F_j^1, \cdots, F_j^K\} \right\}_{j=1}^2$ *is a set of composite principal components of* $\{\xi_1^1, \cdots, \xi_1^N; \xi_2^1, \cdots, \xi_2^N; \}$*. if and only if there exist row vectors* $\mathbf{e}_1, \cdots, \mathbf{e}_K$ *in* $\mathbb{R}^N$*,* $\mathbf{e}_k = (e_k^1, \cdots, e_k^N)$*, such that*

$$\mathbf{e}_k^i \mathbf{C}_i = \lambda_k^i \, \mathbf{e}_k^i, \quad \mathbf{e}_k^m \cdot \mathbf{e}_k^n = \delta_{mn}, \quad F_j^k = \frac{1}{\sqrt{\lambda_k}} \sum_{k=1}^{2} \sum_{i=1}^{N} \xi_j^i \, \mathbf{e}_j^i(k) \quad X_j = \sum_{i=1}^{N} F_1^i f_j^i / \sqrt{\eta^i},$$
$$\forall\, k, l = 1, \cdots, K.$$

In addition,

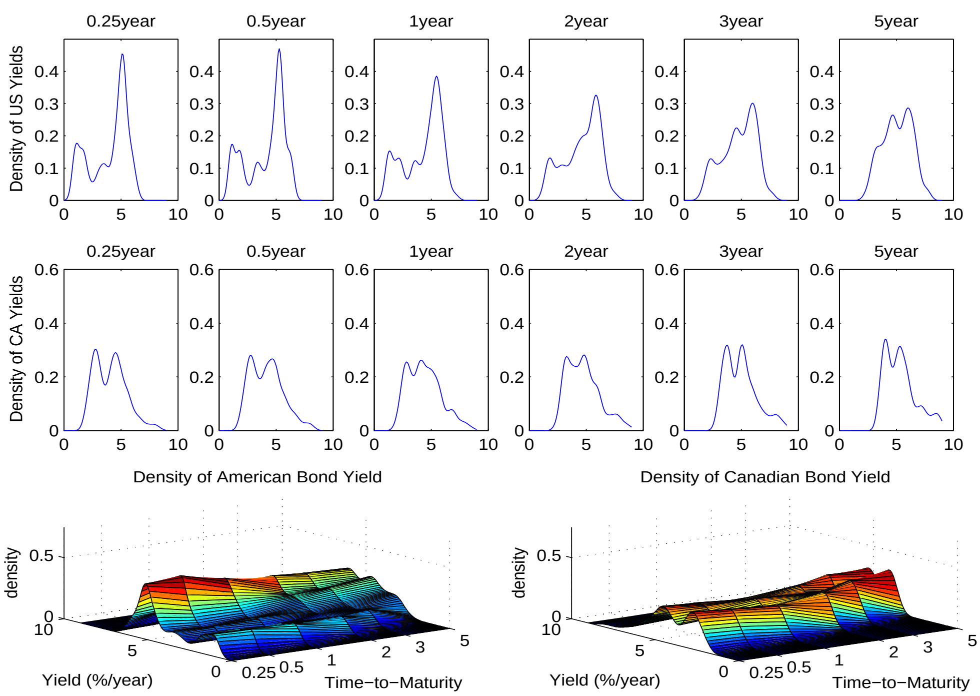
$$\min_{\dim(V)=n} \sum_{i=1}^{N} \operatorname{dist}^2(F^i, V) = \sum_{i=1}^{N} \left\| F^i - \sum_{k=1}^{n} \langle F^i, F^k \rangle X^k \right\|^2 = \sum_{k=n+1}^{N} \eta_k \quad \forall\, n = 1, \cdots, K,$$

$$\min_{\dim(W)=n} \sum_{j=1}^{2} \mathrm{dist}^2(\xi_j^i, W) = \sum_{j=1}^{2} \left\| \xi^i - \sum_{k=1}^{n} \langle \xi^i, F_1^k \rangle F_1^k \right\|^2 = \sum_{k=n+1}^{N} \lambda_k \quad \forall n = 1, \cdots, K,$$

$$R_n := \frac{\sum_{i=1}^N \mathrm{dist}^2(F^i, \{X^1, \cdots, X^n\})}{\sum_{i=1}^N \|F^i\|^2} = \frac{\sum_{k=n+1}^N \eta_k}{\sum_{k=1}^N \eta_k}. \tag{3.3.21}$$

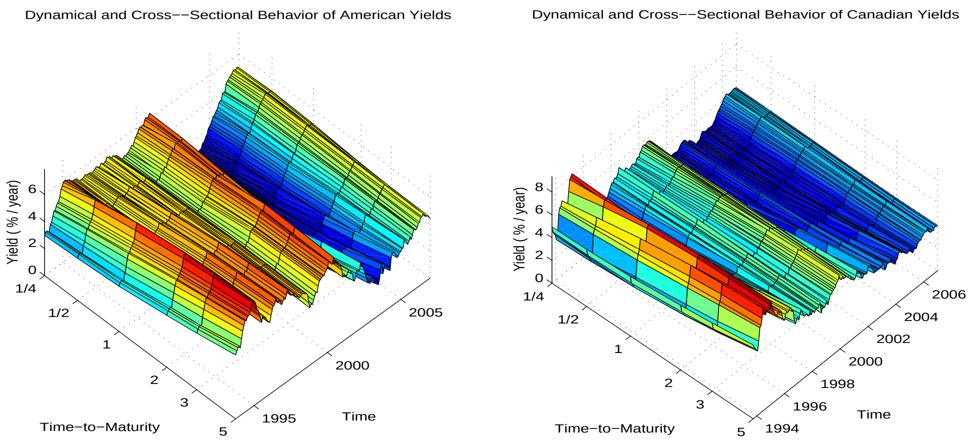
The proof is given in the Appendix A.7. Principal component analysis can be referred by [89] and references therein.

Figure 11: Empirical Densities of Distribution of Yields of Various US and CA Bonds



3.3.7.2 Economic Meaning of Factors and Loads The factors and loads we derived above have economic meaning associated with the macroeconomic variables. The global factors we derived in this paper are the same for both currencies, and

Figure 12: Fixed-Term Time-to-Maturity of US and Canadian Government Bonds



I point out that only the first two factors are used in the empirical work, with the asymmetric effects from the loadings associated with each factor for both countries.

We know that the macroeconomic variables can be combined with yield factors. People can incorporate macroeconomic variables into yield-curve models for the fundamental determinants of interest rates. Diebold, Piazzesi and Rudebusch (2005) discuss how macroeconomic variables should be combined with yield factors. Diebold, Rudebusch and Aruoba (2005) point out that three latent factors (essentially level, slope, and curvature) from a set of yields on U.S. Treasury securities. These factors can be related to three observable macroeconomic variables (specifically, inflation, real activity, and a monetary-policy instrument). Furthermore, they examine the correlations between Nelson-Siegel yield factors and macroeconomic variables. They find that the level factor is highly correlated with the first one, inflation, and the slope factor is highly correlated with the second one, real activity. The curvature factor appears to be unrelated to any of the main macroeconomic variables. In this paper, there is no space to talk about the association between the macroeconomic variables and yield factors, but this is interesting for future research.

3.4 MODELING THE EXCHANGE RATE

3.4.1 The data

Based on the above theoretical framework, in this section we use historical data of US Government bonds rate of various maturities to find a complete Linear Term Structure Model (LTSM). Our data source includes the Economics research Division of Federal Reserve Bank, Bank of Canada, "EconoMagic.com", IMF International Financial Statistics, Canada Statistics and Bureau of Economics Analysis. All our data used are daily adjusted and have been computed to be zero-coupon bond rates. The Canadian-U.S. dollar exchange rates are end of the day markets rates and all variables are in natural logs. c monetary approaches, to maximize the number of time-to-maturity of bonds with available historical data, 3313 complete sets of daily data from 10/1/1993 to 12/29/2006 for 10 different bonds are used in this paper, with time-to-maturities 3-month, 6-month, 1-year, 3-year, 5-year, respectively. The time window is from 10/1/1993 to 12/29/2006. On the other hand, the forward exchange rates data are used in this paper for one-month and three-month ones from 12/31/1996 to 3/6/2009. The compounded one-month and three-month eurocurrency interest rates are also available from 1/1/1975 to 6/12/2009 in this paper to get more empirical backup for the Linear Term Structure Model [6].

Figure 11 lists the empirical densities of bond yields of different terms for each country. It is quite clear that the different terms of bond yields are not normally distributed: each empirical kurtosis is well-below 3, which is of normal distribution. And the short-term yields have statistically significant negative skewness whereas long-term yields have positive skewness.

Figure 12 illustrates the dynamical and cross-sectional behavior of the yields curves for both US and Canadian currencies.The yield-to-maturity curves at any current time t provide an overview for transition of different terms of yields through-

[6] Here I'm truly grateful to Craig Burnside, who shares this valuable data with me.

out the time window. In general, long term interest rates are higher than short term rates, although occasionally the reverse occurs. The dynamical behavior provides valuable information for the statistical investigation of factors.

3.4.2 Interest Rate Differential Examination

To model the exchange rate, the LTSM is a key in this paper from the view of asset pricing. How well the LTSM fits the empirical data, and, in other words, the theoretical derived fitted data matches the empirical actual data is a first job for interest rate pricing. As long as getting a good fitted values for interest rate differentials, the forward premium can also get a good pricing. Finally, the anomaly and the expected excess returns will be examined in the context of the Linear Term Structure Model.

First I want to use the component principal component analysis or global factor analysis in section 3.3.7 finding loads $\ell_k(s)$ and global factors F_t^k in the Composite Principal Component models. [7]

$$r_t^{t+s} = \ell_1(s)F_t^1 + \ell_2(s)F_t^2 \quad \forall\, t \in \mathbf{T},\ s \in \mathbf{S}, \tag{CPCM}$$

where $\mathbf{S} = \{1/4, 1/2, 1, 2, 3, 5\}$ is the list of time-to-maturity for both country and $\mathbf{T} = \{t_i\}_{i=1}^{3313}$ is the historical trading dates.

According to the theorem I proved in section 3.3.7.1, there are two steps to find the global factors for both currencies. First, the eigenvalues of variance matrix of six common factors are derived by the vectors set up with common factors of each term or time-to-maturity yields for both currencies, and they are the basis for formulas to get the global factors. In Table 4, the panel (b) shows the eigenvalues and cumulative

[7] The component principal component analysis or global factor analysis posted in this paper is an alternative way to find the global factors in multi-country environment. Especially, the first step finding the common factors is closely related to macroeconomic variables. Reader can check that three factors will be needed to capture the same contribution of factors if putting all twelve yields together disorderly.

proportions of factors for each term. There are at most two common factors for both currencies with time-to-maturities 3-month, 6-month, 1-year, 3-year, 5-year, respectively. So there are two columns to correspond to each different term.

Figure 3.4.2 also plots out the time series of yields for each term, with the common factors capture the common trend of both US and Canadian yields. The blue lines are the US yield rates, and the red lines are the Canadian yield rates. The green lines are the common factors for different terms. The first common factors counts for about 90% of the cumulative contributions, and will be used for getting the global factors. The loadings attached to the common factors are reported for each currencies. Second, the global factors are derived based on the six common factors. Panel (a) tells us that the first two global factors can explain 99.72% of the total variance of the common factors' matrix. Only the loadings for the first two global factors are needed to report here, and the Two-Factor Composite Principal Component Models are chosen.

Corresponding to the economic meaning of the global factors, the loadings of the first factor are uniformly positive, which means that for every term the yields get higher as the first factors go up. In [81], The first factor is referred to as the *level factor*. The loads of the second factors go from positive to negative monotonically, which means an increment of the second factors rotates the yield-to-maturity curve, so the second factor is called the *slope factor*. Similarly, the third factor is called the *curvature factor*. And in the previous literature, these three factors are closely related to the inflation, real activity, and the monetary policies.

3.4.3 Empirical Two-Factor Models

In this section, I will use the two global factors which are derived above to get the complete sets of variables in the LTSM. For US, on the top of Table 5, American bond empirical loadings and theoretical Loadings are listed for each different term, $1/4, 1/2, 1, 2, 3, 5$ year. For instance, the empirical loadings for 3-month yields are

3.9525 and -0.6661 for the first and second global factor. They are calculated based on the real data by using global factor analysis above. And the 3.9866 and -0.6873 are the theoretical loadings derived from the Riccati equations which are proved in Proposition 1. Similarly, all the rest American bond empirical loadings and theoretical loadings are listed for each different term, so the completer sets of loadings are given corresponding to two global factors. The error and relative error are 0.0052 and 0.0022.

On the other hand, Canadian bond empirical loadings and theoretical loadings are listed on the bottom of Table 5 for each different term, $1/4, 1/2, 1, 2, 3, 5$ year. The empirical loadings for 3-month yields are 4.0664 and -0.2337 for the first and second global factor. And the 4.0638 and -0.2498 are the theoretical loadings derived from the Riccati equations. Similarly, all the rest Canadian bond empirical loadings and theoretical loadings are listed for each different term, so the completer sets of loadings are also given corresponding to the same two global factors.

The error and relative error are 0.0023 and 0.0009. There is an important view here to see. The response, or the size of the loadings for US currency is slightly smaller than the Canadian currency consistently for each term of first global factor, or level factor. The response of the loadings for each term of second global factor, or curvature factor needs to be examined in details. The size of the loadings ranges from -0.6873 to 0.3170 for US, but Canada gets -0.2498 to 0.7189. US loadings have greater response for short-term yields, but Canadian loadings have greater response for long-term yields. Especially, for 1-year and 3-year bonds, they have big difference. Monotonically, the turning point of US yield loadings is at 3-year term, but Canadian yield loadings change the sign from negative to positive at 2-year term. This tells us the asymmetric effect of the response of each currency's loadings is mostly from the curvature factor.

To see the effectiveness of the two-factor linear term structure model, I need to check the relative errors compared with the values of the interest rates of the yield-to-maturity curves. As shown in Tables 5, the average error for yields of US currencies is 0.0167%, corresponding to the average values of US interest rates of the yield-to-maturity curves about 4-5%. And the 0.0195% is attained for the Canadian values, about 3-5%, of the interest rates of the yield-to-maturity curves. In summary, two-factor empirical models are obtained for both currencies. In Figure 16 and 17 we illustrated the accuracy of the LTSM. Dots are empirical data, solid curve is the LTSM. The first two are best and worst fits. The dates of the remaining 7 plots are randomly picked. From the figures and Table 5, one can conclude that the two-factor models fit the historical data quite well, without imposing the symmetric restriction of factors on yields.

After checking the the effectiveness of LTSM, I also plot out the curves for time-to-maturity theoretical loads. The fitness between the loads from empirical data via the CPCM and the loads from solutions of Riccati equations is displayed in Figure 14 and Figure 15. In both Figures, two time-to-maturity scales are used: one is the standard scale; the other is the $\log[1/4+s]$ scale for $s \in [0, 30]$, where actual values of s are marked. Using both scales, we can see the overall fit of the theoretical LTSM to the empirical data. In the figure, we plot both the loads $L_i(s)$, $i = 1, 2$, in the bond price formula $-\log B_t^{t+s} = L_1(s)F_t^1 + L_2(s)F_t^2$ and the loads $\ell_i(s) = L_i(s)/s$ in the yield formula $r_t^{t+s} = \ell_1(s)F_t^1 + \ell_2(s)F_t^2$. The load $L_0(s)/s$ represents the mean yield. Using the solution $(L_1(s), L_2(s))$ of the Riccati equations (3.3.6) and the (rotated) factors (F_t^1, F_t^2) of CPCM, we then obtain the complete description of the LTSM model. Following a similar procedure, the fit between theoretical loads obtained by solving (3.3.6) and those from empirical data is checked for Canadian currencies too. Note that the first and second global factor are plotted out too, and keep in mind that they are exact the same for both time-to-maturity yield curves.

3.4.4 LTSM Guarantee Positive Interest Rates with Suitable State Space

According to the Theorem stated in Chen and Huang [24] (2008), I have Linear Term Structure models with good state spaces, in which the Black-Scholes equation admits a unique solution and the sample path $\{(F_t^1, F_t^2) \mid t \in \mathbf{T}\}$ contains. For US currency, the sample path $\{(F_t^1, F_t^2)\}_{t\in\mathbf{T}}$ is obtained through the numerical values which are in a vicinity of those empirical ones calculated from the global factor analysis. After solving the Riccati equations, the optimal parameters for the solution come out matching the empirical ones from CPCM. Similarly, for Canadian currency, the sample path is exactly the same as the path of US currency, because the two global factors are the same for both currencies. And the optimal parameters for the solution matching the empirical data for Canadian currency are also calculated. The details to set up the boundary or the lines for each term in Figure 18 will not be illustrated in this paper, and they can be searched in Chen and Huang [24] (2008).

Figure 18 (a) and (b) are the state space figures for US and Canadian yields, under which the variances matrix are linear structure of factors and the yields are always positive. It shows the boundary of state space in which $\sigma(z) > 0$ can be seen from the upper and lower lines which are closest to the trajectories of the global factors. Furthermore, for each $s \in (0, 5]$, $(L_1(s), L_2(s)) \cdot z > 0$ for LTSM. Each thin line is given by the equation $(L_1(s^i), L_2(s^i)) \cdot z = 0$ for LTSM, $i = 1, \cdots, 5$. Also $\sigma(z)\mathbf{n}(z) = 0$ for $z \in \partial\Omega$. The Brownian motion like trajectory is the sample path $\{(F_t^1, F_t^2) \mid t \in \mathbf{T}\}$ for LTSM which stays in state space. Figure 18 shows that the trajectory will stay in the state space in which the interest rates are always positive. The results for expected rate of depreciation, forward premium, and risk premium will further give the positive support for the models.

3.4.5 Forward Premium, Risk Premium and Anomaly Equation

Forward premium and risk premium are to be explored in this section. The anomaly equation will be examined by the factors which are used in the environment of LTSM. The goal is firstly to show the anomaly phenomenon exists not only from the historical data but also from the theoretical fitted values, especially the fitted interest rates differentials, and secondly to show the negative correlation between the forward premium and the expected rate of depreciation resulting from the behaviors of factors used in LTSM.

The expected rate of depreciation and forward premium get checked using the forward exchange rates and the Euro-currency interest rates. First, the Covered Interest Parity and the Uncovered Interest Parity are listed in Table 6. The simple OLS is used to give us a preliminary idea, like previous literature. The results from part (a) and (b) show that CIP works fine but UIP does not, which supports the traditional view for CIP and UIP. Second, the forward premium anomaly equations are examined both by the forward rates themselves and the term structure models. We can tell that the term structure models are doing well, so the expected rate of depreciation and forward premium get new support in the context of LTSM. This can be apparently told as long as the empirical fits of LTSM to actual yield curves get evidence for both currencies.

The forward premium and the anomaly equation are examined again directly by the factors derived above and used in LTSM, and I check whether it's consistent with Fama's conditions. From Table 7, depreciation rates of real data are showed, and there are big correlations within the time series. The forward premium is shown by the interest differentials. Forward premium approximations are carried out on the two global factors. F_t^1 is the level factor, and the loadings associated with it are close to zero. F_t^2 is the slope factor, and the loadings are all negative. On the other hand, expected depreciation rates are regressed on the two global factors. The loadings associated with the level factor are significantly negative, which is quite different from the forward premium regression, which is consistent with the theoretical derivation in previous section. The loadings for the slope factor are small, which is apparently smaller than those of forward premium approximations. A possible reason for the significant asymmetry response to factor processes is that the money demand in the market is an information asymmetric procedure. Two factors are capable of accounting for the systematic risks and idiosyncratic risks. CIP do hold but UIP does not for the extra K-risk issue. For the Fama's conditions checking, which is consistent with the previous researches. Condition (1), $\mathrm{Cov}(p, q) < 0$, ranges from -0.0010 to -0.0012 for different time-to-maturity. Second, condition (2), $\mathrm{Var}(p) > \mathrm{Var}(q)$, gets proved by the linear combinations of global factors. Third, the derived expected rate of depreciation and forward premium get new support in the context of LTSM. Moreover, the results from Table 7 illustrate a new insight for risky environment. Finally, in part (f), Fama's conditions are checked again consistent with the previous researches.

In summary, Figure 19 makes a promising picture for risk premium, forward premium and expected rate of depreciation for different time-to-maturity of yields. The blue lines with dots are risk premium for different time-to-maturity of yields. The green lines are expected rate of depreciation for different time-to-maturity of yields. The $p_t = f_t - \mathbb{E} \log S_{t+1}$ represents the risk premium, and $q_t = \mathbb{E} \log S_{t+1} - \log S_t$ represents the expected rate of depreciation. The red lines are sum of risk premium and expected rate of depreciation for different time-to-maturity of yields, that is, $p_t + q_t$, forward premium. They have consistently negative correlation between each other,except for the 5-year yields. The empirical results do give big support on the validity of LTSM as a model of exchange rate determination, Moreover, the anomaly equation get more examinations with the LTSM.

3.5 CONCLUSION

The empirical findings for the exchange rate dynamics incorporated with LTSM from this paper are consistent with some of the previous finding as many researchers have achieved for data series from 1978. This paper shows that the LTSM explain the movements of the Canadian dollar - US dollar exchange rate from 1990 to 2008. A possible reason for the significant asymmetry response to factor processes is that the money demand in the market is an information asymmetric procedure. Two factors are capable of accounting for the systematic risks and idiosyncratic risks. But The empirical results do give big support on the validity of LTSM as a model of exchange rate determination, especially for the forward premium.

The major findings in this paper are: first, the empirical results give support on the validity of LTSM as a model of exchange rate determination through Riccati solutions. With the theoretical loads obtained from the solutions of the ordinary differential equations of Riccati type, the resulting bond price B_t^T has the property that $B_t^T < 1$ for $T \in (t, t+5]$. Second, the result of state space restriction further backup the LTSM in forecasting in the exchange rate in the long run. The state space Ω and parameters $(\sigma_k^{ij}, p_k^i, c_k)$ are chosen such that the Black-Scholes partial differential equation for pricing is completely determined and has a unique solution for every bounded payoff function. Third, there is significant evidence found in testing the speculative bubble, which shows that the state variables or stochastic factors are capable of accounting for the systematic risks and idiosyncratic risks. Forth, the LTSM model suggested some evidence for the short-run dynamics, and the interest rate terms are compatible of exchange rate terms, which shows that the LTSM model is a good estimate for the forecasting in the short-run movement for the exchange rate. The empirical sample path stays in Ω. The yield surface produced from the LTSM matches the empirical one at most points. Thus, we have two term structure models that are simple, consistent, and accurate. Finally, the expected rate of depreciation and the risk premium get new derivations in this paper, with the empirical backup. There is some evidence in the interrelationship between the expected rate of depreciation and the forward premium, and fluctuation of the risk premium is shown to be greater than the forward premium in the context of LTSM in this paper.

Table 4: The Two-Factor Composite Principal Component Model

Panel (a) Cumulative Contributions of Global Factors

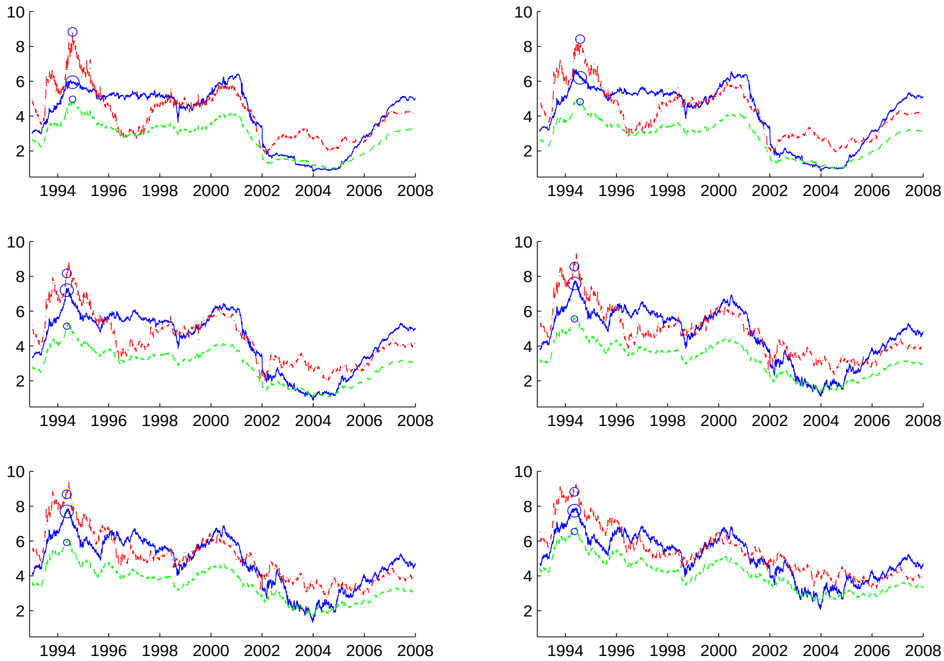
Factors k	1	2	3	4	5	6
	The eigenvalue of variance matrix of six global factors [1]					
λ_k	5.6886	0.2947	0.0150	0.0013	0.0002	0.0001
Cumulative Proportion	0.9481	0.9972	0.9997	0.9999	1.0000	1.0000
	Loadings for the first two global factors [2]					
Maturity (Year)	1/4	1/2	1	2	3	5
Loadings for the first global factor	0.1686	0.1715	0.1747	0.1753	0.1726	0.1643
Loadings for the second global factor	0.9339	0.7449	0.3324	-0.2349	-0.6155	-1.1922

Panel (b) The first step to get the common factors of American and Canadian bonds [3]

Maturity (Year)	1/4		1/2		1		2		3		5	
	The eigenvalue of variance matrix and the first common factors of American and Canadian bonds											
λ_k	4.1479	0.5567	4.3670	0.5212	4.3976	0.4776	4.1760	0.4296	3.8091	0.3817	3.2171	0.3026
Proportion	0.8817	1.0000	0.8934	1.0000	0.9020	1.0000	0.9067	1.0000	0.9089	1.0000	0.9140	1.0000
Loadings	0.3866	0.8262	0.3756	-0.8581	0.3653	0.9300	0.3678	1.0063	0.3725	1.1113	0.3758	1.3429
	0.3027	-1.0552	0.2965	1.0873	0.3065	-1.1085	0.3227	-1.1468	0.3518	-1.1768	0.4119	-1.2252

1. The eigenvalues of variance matrix of six common factors are derived by the vectors set up with common factors of each term or time-to-maturity yields for both currencies, and they are the basis for formulas to get the global factors. 2. Only the loadings for the first two global factors are reported here, since the Two-Factor Composite Principal Component Models are chosen. 3. In the first step, the first common factors are derived for each term or time-to-maturity yields for both currencies.

Figure 13: Two Country First-Step Common Factor for 1/4, 1/2, 1, 2, 3, 5 year yield rates.

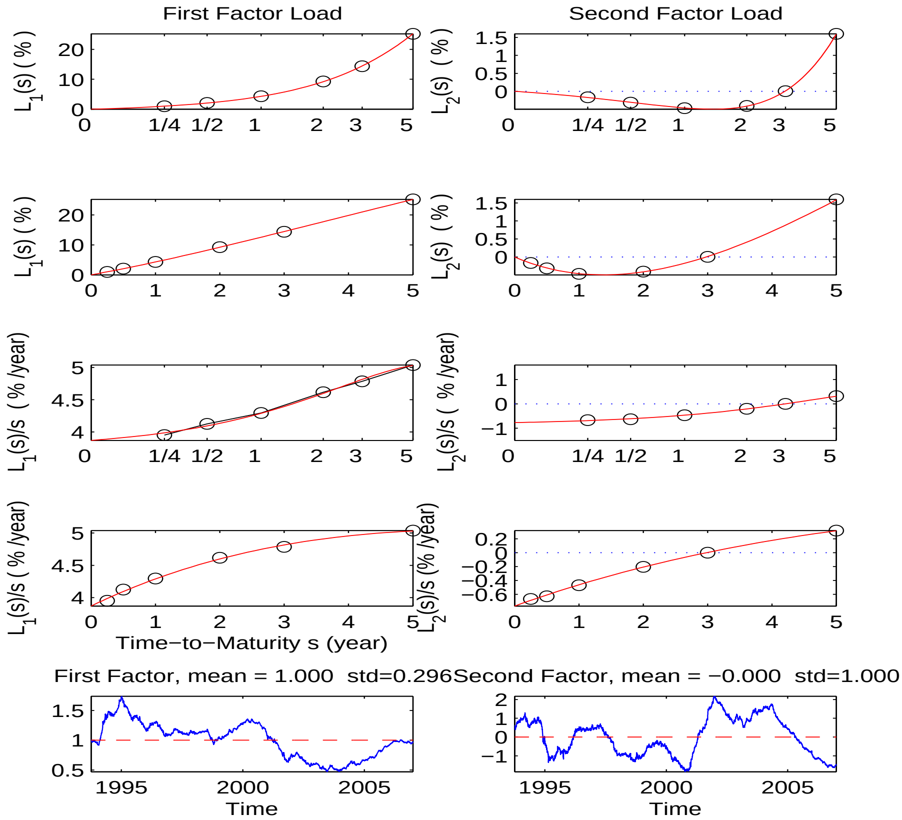


The blue lines with the biggest circle are the US yield rates, and the red lines with medium circles are the Canadian yield rates. The green lines with the smallest circles are the common factors for different terms. Dots are used for clarifying the common factors in case of no colors.

Table 5: Effectiveness of the Two-Factor Linear Term Structure Model

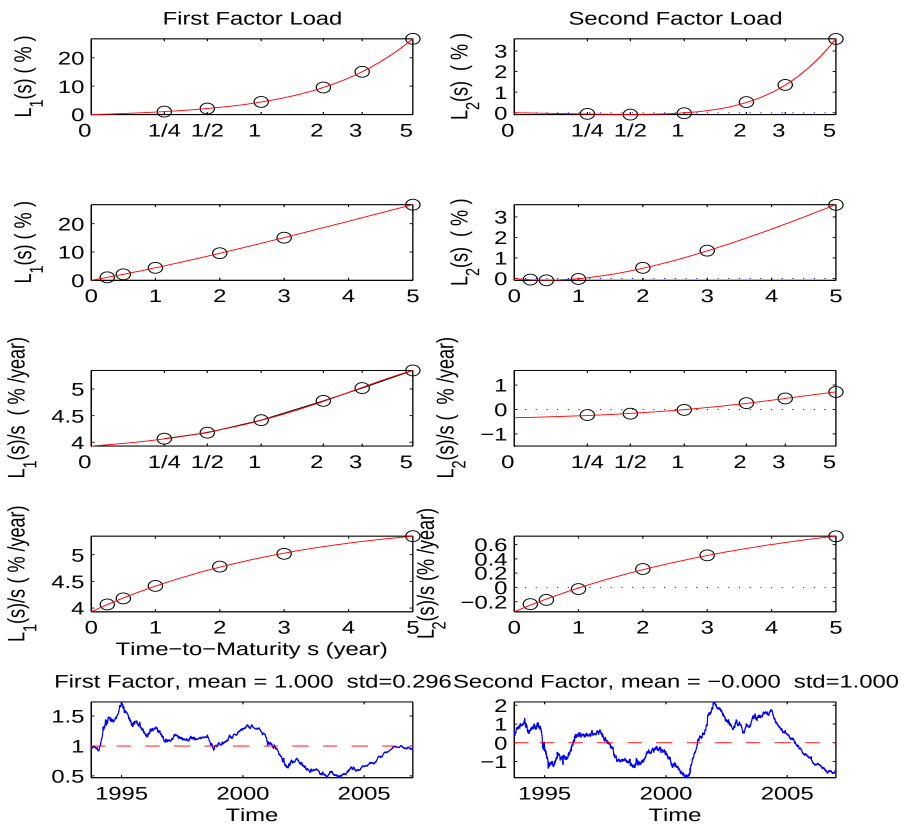
	1/4 year	1/2 year	1 year	2 year	3 year	5 year
(a) American Bond Empirical Loadings and Theoretical Loadings						
Theoretical Loading	3.9866	4.0950	4.2897	4.5991	4.8162	5.0314
	-0.6873	-0.6091	-0.4627	-0.2072	0.0042	0.3170
Empirical Loading	3.9525	4.1261	4.2960	4.6177	4.7863	5.0392
	-0.6661	-0.6288	-0.4678	-0.2038	0.0015	0.3201
Error	0.0052					
Relative Error	0.0022					
(b) Canadian Bond Empirical Loadings and Theoretical Loadings						
Theoretical Loading	4.0638	4.1879	4.4104	4.7673	5.0292	5.3463
	-0.2498	-0.1642	-0.0087	0.2482	0.4470	0.7189
Empirical Loading	4.0664	4.1796	4.4148	4.7773	5.0177	5.3490
	-0.2337	-0.1756	-0.0236	0.2574	0.4506	0.7163
Error	0.0023					
Relative Error	0.0009					
(c) Average Errors for yields of both currencies	US	0.0167%		Canada	0.0195%	

Figure 14: Fitness Between Empirical and Theoretical Loads of LTSM for American Bonds



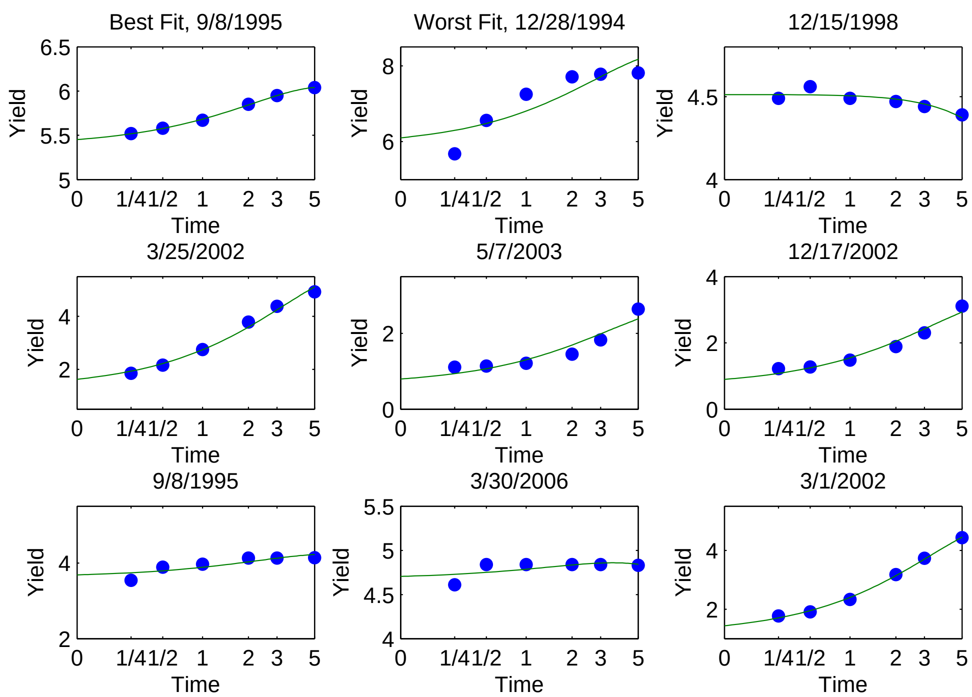
The LTSM loads are obtained by solving the Riccati equations (2.2.6). The induced mean is obtained by the formula $L_0 = L_1\text{mean}(F^1) + L_2\text{mean}(F^2)$. CPCM loads are obtained from the Principal Component Model.

Figure 15: Fitness Between Empirical and Theoretical Loads of LTSM for Canadian Bonds



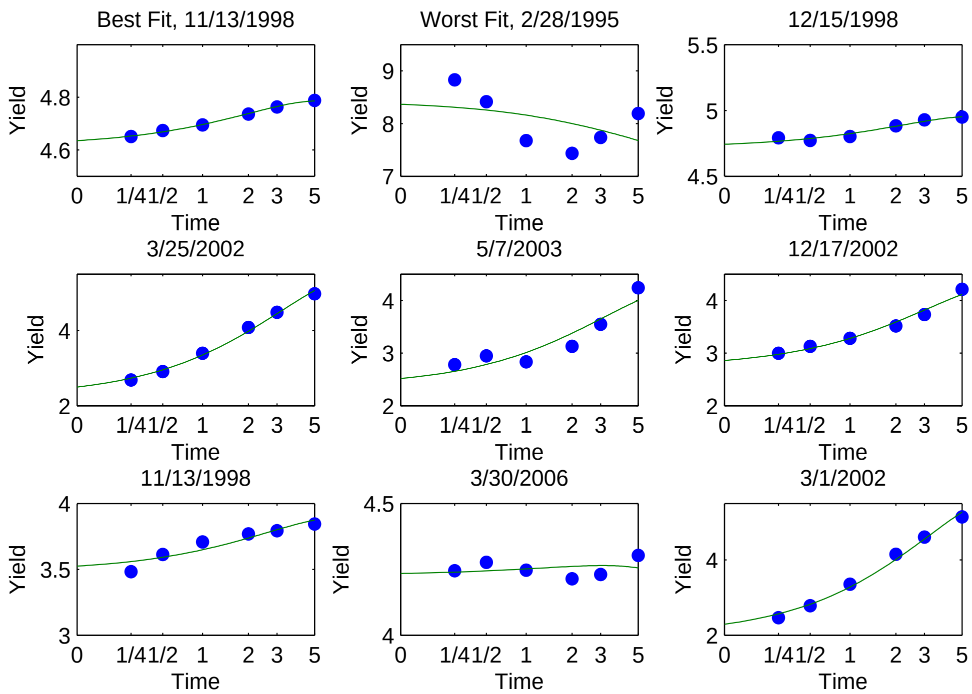
The LTSM loads are obtained by solving the Ricartti equation (2.2.6). The load $L_0(s)/s$ represents the mean yield. CPCM loads are obtained from the Principal Components Model.

Figure 16: Empirical Fits of LTSM to American Yields



Dots are empirical data, solid curve is the LTSM. The first two are best and worst fits. The dates of the remaining 7 plots are randomly picked.

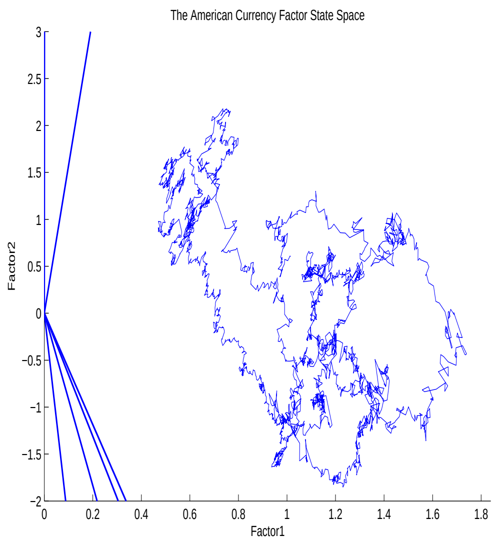
Figure 17: Empirical Fits of LTSM to Canadian Yields



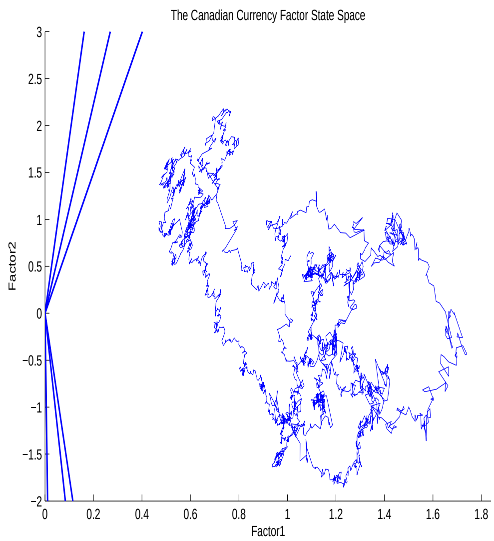
Dots are empirical data, solid curve is the LTSM. The first two are best and worst fits. The dates of the remaining 7 plots are randomly picked.

Figure 18: The State Space Ω

(a) American Bonds



(b) Canadian Bonds

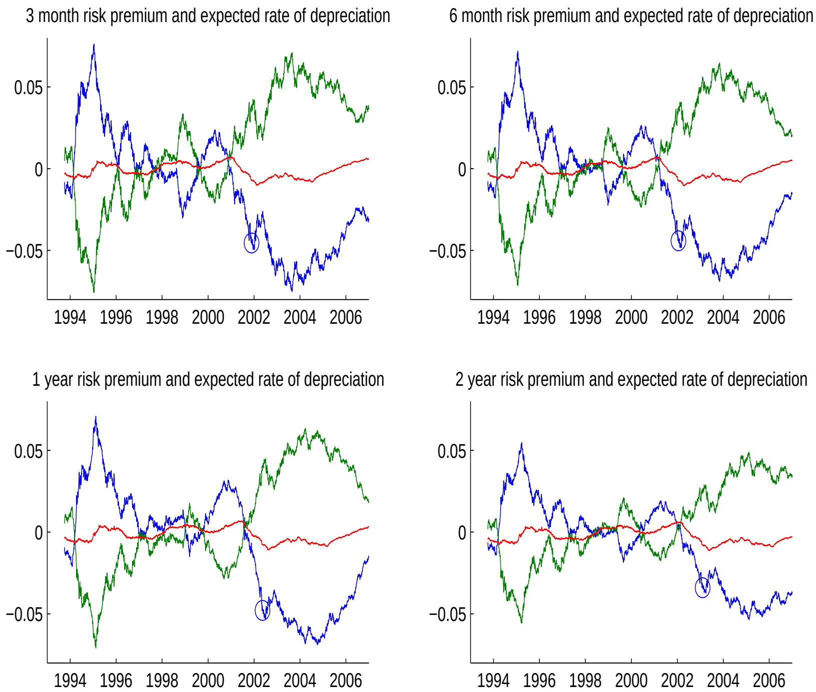


The boundary of Ω in which $\sigma(z) > 0$ can be seen from the upper and lower lines which are closest to the trajectories of the global factors. Furthermore, for each $s \in (0, 5]$, $(L_1(s), L_2(s)) \cdot z > 0$ for LTSM Each thin line is given by the equation $(L_1(s^i), L_2(s^i)) \cdot z = 0$ for LTSM $i = 1, \cdots, 5$. Also $\sigma(z)\mathbf{n}(z) = 0$ for $z \in \partial\Omega$. The Brownian motion like trajectory is the sample path $\{(F_t^1, F_t^2) \mid t \in \mathbf{T}\}$ for LTSM which stays in Ω.

Table 6: Forward Premium Anomaly Examination with Forward Data

Currency	Term	1 month	3 month
(a) Covered Interest Parity			
$f_t - s_t = r_t^T - r_t^{T*}$	a_1	1.0338	1.0088
		(0.0038)	(0.0022)
(b) Uncovered Interest Parity			
$s_T - s_t = r_t^T - r_t^{T*}$	a_1	-2.2806	-1.5858
		(0.4365)	(0.2616)
(c) Forward Premium Regressions Using Data			
$s_T - s_t = a_0 + a_1(f_t - s_t) + \varepsilon_t$	a_0	0.0452	0.0242
		(0.0048)	(0.0029)
	a_1	-2.5878	-2.0788
		(0.4145)	(0.2615)
(d) Forward Premium Regressions			
$s_T - s_t = a_0 + a_1(r_t^T - r_t^{T*}) + \varepsilon_t$	a_0	0.0454	0.0236
		(0.0047)	(0.0029)
	a_1	-2.9759	-1.9779
		(0.4364)	(0.2631)

Figure 19: Risk Premium and Expected Rate of Depreciation for Different Term of Yields



The blue lines with dots are risk premium for different time-to-maturity of yields. The green lines are expected rate of depreciation for different time-to-maturity of yields. The $p_t = f_t - \mathbb{E} \log S_{t+1}$ represents the risk premium, and $q_t = \mathbb{E} \log S_{t+1} - \log S_t$ represents the expected rate of depreciation. The red lines are sum of risk premium and expected rate of depreciation for different time-to-maturity of yields, that is, $p_t + q_t$, forward premium.

Table 7: Descriptions and Estimation Results for LTSM on Forward Premium Anomaly

Currency	Term	3 month	6 month	1 year	2 year	3 year	5 year
(a) Depreciation Rate							
$s_{t+1} - s_t$	Mean	0.0016	0.0031	0.0058	0.0116	0.0148	-0.0106
	Var	0.0008	0.0017	0.0035	0.0088	0.0145	0.0182
	Std	0.0286	0.0415	0.0592	0.0937	0.1202	0.1350
	Corr	0.9817	0.9910	0.9956	0.9983	0.9991	0.9992
(b) Forward Premium							
$f_t - s_t = r_t - r_t^*$	Mean	-0.0004	-0.0005	-0.0020	-0.0057	-0.0102	-0.0152
	Var	0.0000	0.0000	0.0001	0.0004	0.0007	0.0015
	Std	0.0028	0.0054	0.0100	0.0190	0.0266	0.0390
	Corr	0.9959	0.9972	0.9961	0.9945	0.9939	0.9927
(c) Forward Premium Approximations $f_t - s_t = (L_1 - L_1^*)F_t^1 + (L_2 - L_2^*)F_t^2$							
	$L_1 - L_1^*$	-0.0772	-0.0929	-0.1207	-0.1682	-0.2130	-0.3149
	$L_2 - L_2^*$	-0.4375	-0.4449	-0.4540	-0.4554	-0.4428	-0.4019
(d) Expected Depreciation Rate Regressions $s_{t+1} - s_t = a_0 + a_1 F_t^1 + a_2 F_t^2 + \varepsilon$							
	a_0	0.1414	0.1254	0.1159	0.0978	0.1032	0.0616
		(0.0085)	(0.0061)	(0.0040)	(0.0029)	(0.0022)	(0.0026)
	a_1	-0.1277	-0.1151	-0.1079	-0.0901	-0.0925	-0.0545
		(0.0082)	(0.0059)	(0.0038)	(0.0027)	(0.0020)	(0.0023)
	a_2	-0.0118	-0.0076	-0.0020	0.0076	0.0083	-0.0030
		(0.0025)	(0.0018)	(0.0012)	(0.0009)	(0.0006)	(0.0005)
(f) Fama's Condition Check							
Condition (1) $\mathrm{Cov}(p, q) < 0$		-0.0011	-0.0010	-0.0011	-0.0006	-0.0012	0.0001
Condition (2) $\mathrm{Var}(p) > \mathrm{Var}(q)$	$\mathrm{Var}(p)$	0.0012	0.0011	0.0012	0.0007	0.0013	0.0005
	$\mathrm{Var}(q)$	0.0011	0.0010	0.0010	0.0006	0.0011	0.0001

4.0 AN INVESTIGATION OF NEW INTERNET MEASUREMENT ON INTERNATIONAL TRADE

4.1 INTRODUCTION

During the past decade there has been much discussion in the popular press about the role that the Internet has had in expanding commercial activity, possibly including international trade. In several papers, Freund and Weinhold (2000, 2002 & 2004) find evidence that the Internet stimulates international trade. To measure Internet development across countries they use "data from the Internet Software Consortium on the number of web hosts attributed to each country that is obtained from counting top–level host domain names"[8]. As they point out, however, this may not be a very good measure, since there is no necessary correlation between a host's domain name and where the site is actually located. In this paper, I develop several new measures of Internet usage and use these data to re-test the relationship between Internet usage and international trade.

This paper carries out the investigation by developing a parallel processing computational statistics method in order to transfer the original Internet measurement data source to the readable dataset under the global comparative environment. In this way, we can study the effects of the Internet cross-traffic traveling by looking at several attributes, including the round trip times (RTT), which characterizes macroscopic connectivity and performance of the Internet and allows various topological

[8]Freund and Weinhold (2004) pg. 171.

and geographical representations at multiple levels of aggregation granularity.

To measure the connectivity and the performance of the Internet, i.e., the efficiency of the information traveling through the hops, we use data from the Cooperative Association for Internet Data Analysis (CAIDA). Utilizing Skitter data [9], we can map dynamic changes in the Internet topologies by tracking related performance effects in Real-Time, using Skitter's RTT data to indicate regions of the infrastructure experiencing abnormal delay, and comparing the Internet graphs across time. The data gathered from Skitter provides a valuable input for empirically modeling of the Internet behavior and properties. The CAIDA–Skitter sends 52-byte Internet Control Message Protocol (ICMP) echo request packets from an IP address belonging to one of its current monitors, which provides complexities in the current and future Internet. In other words, Skitter research supplies a new insight into the complexity of a large, heterogeneous, and dynamic worldwide topology.

With the above original data, we developed a parallel processing computational method to do the statistical process to generate a panel dataset of different countries. Therefore, we found a new way of measurement for the diffusion of the Internet, and are the first to set up a unique and valuable dataset, using a panel of 10 countries examined from 1998 to 2007. Different from the previous literature, this project uses city-level daily databases downloaded from CAIDA to construct the yearly, monthly, and weekly data across countries. In investigating the Internet's influence in the cross-country environment, the determinants of the Internet consist of several attributes which capture the stability and the efficiency of the information traveling and characterize the macroscopic connectivity and performance of the Internet. The dataset includes those from 20 cities around the world during a period of 10 years. This allows me to compare the degrees of change in international trade over a large number of countries before and after the beginning of the century. Under these new

[9] Skitter is a tool for actively probing the Internet in order to analyze topology and performance. The CAIDA performs large-scale topology measurements on Skitter, which employs an improved measurement methodology.

measurements of the Internet, a significant and positive relationship can be found between the Internet distance and the different bilateral trade volume. Furthermore, the magnitude of elasticity is discussed for all four different Internet measurements. The results further support the conclusion of Freund and Weinhold (2000).

4.2 LITERATURE REVIEW

Freund and Weinhold (2000) found that countries with relatively more hosts would trade more, simply because they produce and consume a lot of high-tech products. The data in their paper was from the Internet Software Consortium (ISC). It was used to count how many web hosts were attributed to each country by counting top-level host domain names. A top-level domain name is either an ISO country code or one of the generic domains (com/org/net/etc). However, a host could easily be located in the U.S. or any other country, e.g., hosts under the domains EDU/ORG/NET/COM/INT could be located anywhere. Furthermore, although Freund and Weinhold (2000) also used the number of the Internet users in each country provided in the World Bank to get the ratio on the hosts numbers per capita. Hofstader (2004) [60] illustrated in his paper "Internet Accessibility: Beyond Disability"that the web has become increasingly pervasive as the Internet grows and technologies spread with many devices previously thought of as discrete becoming part of our networked world. This support the necessity of choosing another measurement for the diffusion of the Internet besides the data provided by World Bank. Greenstein and Prince (2006) [54] analyzed the diffusion of the Internet across the United States over the past decade for both households and firms, considering costs and benefits on the demand and supply side and discussed the unequal availability and use of the Internet. Their results are based on the Internet use, cable connection and home adoption within the national division. Goldfarb and Blum (2006) [55] use gravity equations to measure international electronic commerce. They show gravity equations hold in the case of digital goods that are consumed over the Internet and have no trading costs, but the effects of distance on electronic commerce mainly focus on the digital goods, which depend on taste. Ward (1996) [92] examined how the free flow of information affects political activities. They examined political behavior by relying on imperfect information to explain deviations from welfare maximization in the political process to derive a positive impact of the Internet, but then still need the empirical data to support their theoretical findings.

On the other hand, the CAIDA has lately posted a number of research papers using the databases set up by the San Diego supercomputing center, partner of Teragrid network, a infrastructure combining leadership class resources at eleven partner sites to create an integrated, persistent computational resource, which are discussing hardware improvements. For instance, Antonelli and Honeyman (2000) [9] demonstrated that with network security threats and vulnerabilities increasing, solutions based on online detection remain attractive, particularly at public institutions; and in the public's perception. Firoiu, Boudec, Towsley, and Zhang (2002) [46] pointed out recent advances in theories and models for Internet Quality of Service (QoS).

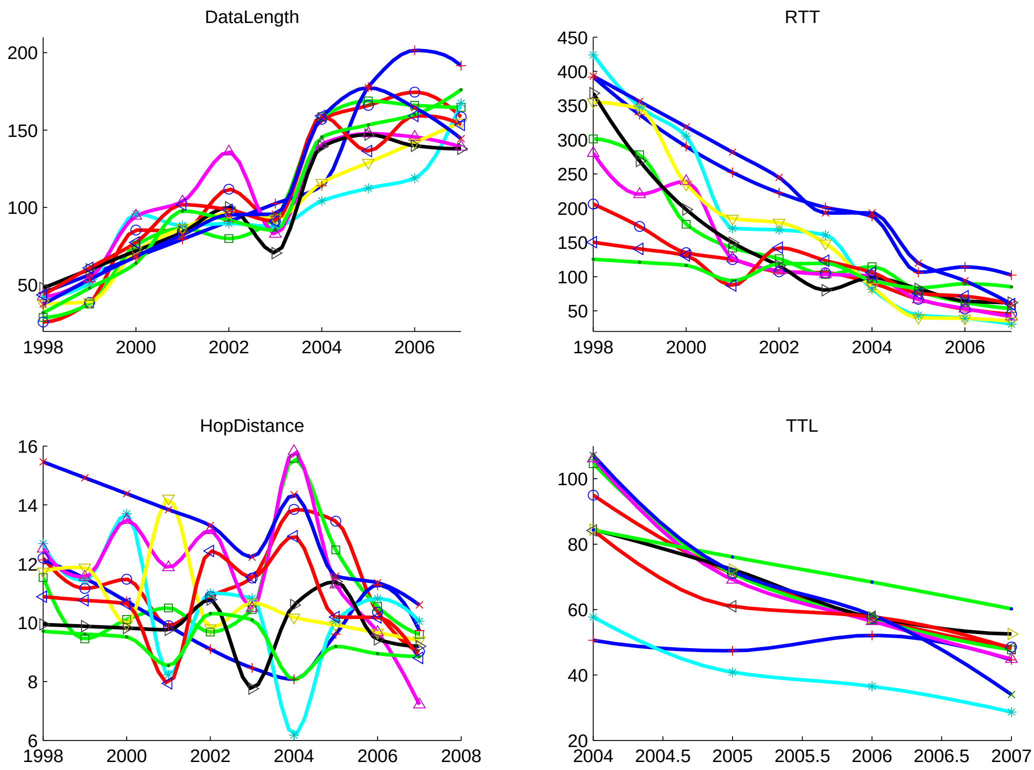
The rest of the paper is organized as follows. Section 3 is devoted to an establishment of the Internet measurement and the computing methodology. Detailed data processes and key computational steps are illustrated. The descriptive results are discussed for both topics. The data is then applied in Section 4 to the gravity equation to see the relationship in the traditional econometric environment, using a panel framework of a cross–sectional times series. The elasticity analysis and moment conditions are discussed in Section 5. In Section 6, we present a dynamic panel causality analysis to demonstrate further empirically that the causality of the Internet diffusion on the bilateral trade volume. Section 7 concludes the paper.

4.3 DATA AND COMPUTING METHODOLOGY

4.3.1 the Internet Data

All data were collected from about ten activated monitors across seven countries for the years 2001-2006. There are seven variables inside our data files, and four of them are the determinants we will discuss in the empirical analysis:

Figure 20: Dynamic Time Plots for Internet Indicators

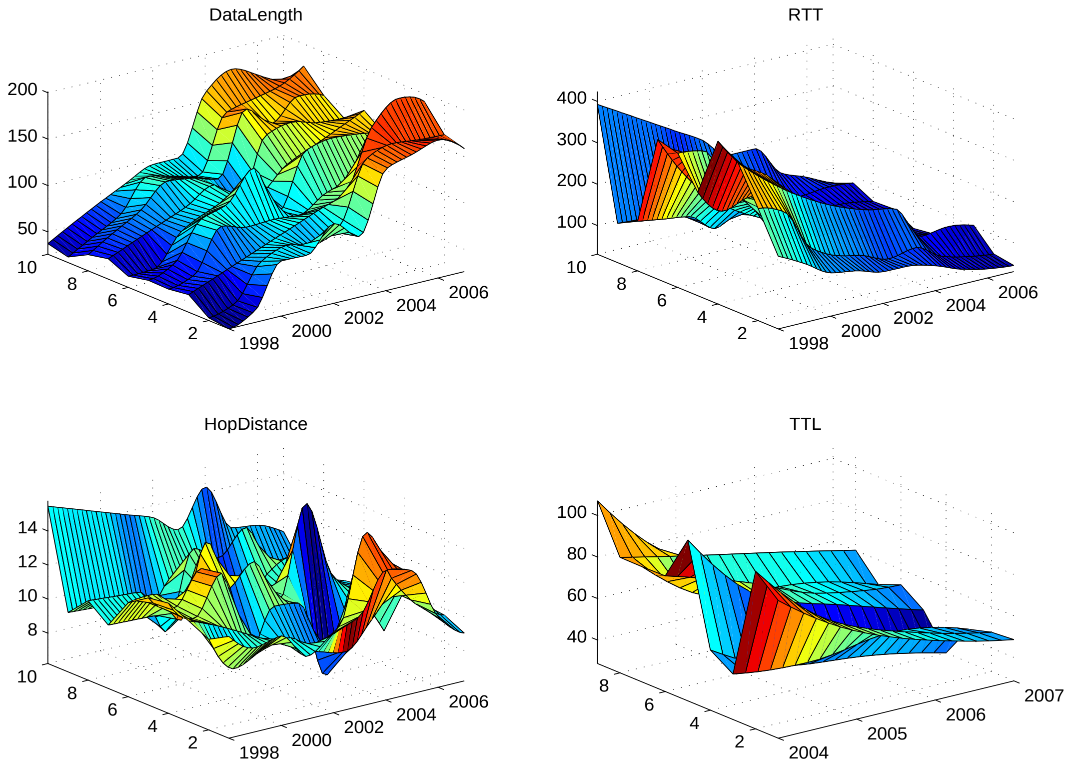


In each subplot, there are ten curves, with each representing the dynamic time plot for different country. The colors and symbols of each curve are denoted as: US(red,○), UK(green,□), CN(blue,+), JP(cyan,∗), CA(magenta,△), KR(yellow,▽), FR(black,▷), SE(red,◁), NL(green,●), NZ(blue,×).

(1) *attr_length*. This variable represents the attributes of the IP address from San

Diego. So if we want to check the location of different cities across countries, we can use this variable to represent different cities in our testable experiments.

Figure 21: Cross-sectional Dynamic Time Series Plane for Internet Indicators



The surface $y = y_t^j$, $1 \leqslant j \leqslant 10, 1998 \leqslant t \leqslant 2007$ of the variables of 10 countries, US, UK, CN, JP, CA, KR, FR, SE, NL, NZ, on 10 different dummies from 1998 to 2007.In the subplot for TTL, the time range is from 2004 to 2007, and the country, KR was excluded.

(2) *data_length*. This represents the length of the signal sent from the source IP address, say, San Diego, to a specific destination IP address. This variable gives us an opportunity to check the efficiency of the information transformation if we change the size of the length or the ratio of the signal length to the number of the hops.

(3) *Src*. This gives the exact figure of the source IP address, say, San Diego. This variable should correspond to the variable attr_length. We can use this to check if

the signal was sent from the same source monitor.

Table 8: The descriptive table of "Data_length" for 10 countries (1998–2007)

Year	US	UK	CN	JP	CA	KR	FR	SE	NL	NZ
1998	26.00	29.00	45.00	42.00	42.60	38.00	48.00	43.49	32.00	37.49
1999	38.69	37.63	56.50	50.82	54.13	39.06	60.00	60.96	48.00	53.46
2000	85.45	75.78	68.00	95.87	94.61	71.87	72.00	77.41	64.00	68.41
2001	84.83	86.63	79.50	88.40	103.77	86.66	84.00	101.88	97.80	82.34
2002	111.80	79.93	91.00	89.67	136.05	96.02	100.10	98.51	91.75	95.25
2003	91.20	93.35	102.50	87.60	82.91	92.92	70.59	91.11	85.26	95.73
2004	156.92	158.65	114.00	104.28	141.13	116.00	139.43	159.14	145.62	158.58
2005	165.86	168.80	177.82	112.42	147.72	129.00	147.09	136.43	153.34	177.36
2006	174.52	165.99	201.45	118.93	145.53	142.00	139.85	159.16	160.00	164.45
2007	158.49	164.63	191.57	167.36	139.15	155.00	137.87	153.32	176.00	144.50
Mean	109.38	106.04	112.73	95.73	108.76	96.65	99.89	108.14	105.38	107.76
Std Dev	53.30	54.25	57.50	35.09	39.22	39.95	38.01	41.92	50.47	49.86
Skewness	−0.22	−0.02	0.52	0.37	−0.55	−0.14	0.07	−0.05	0.02	0.11
Kurtosis	1.73	1.52	1.75	3.18	1.84	1.95	1.37	1.67	1.61	1.56

Last four rows are statistics for measurable variable "Data_length" across time for each country during 1998–2007.

(4) *Dst*. This gives the exact figure of the destination IP address. San Diego sent out the invitation to thousands of IP addresses across the world, and any hops echoing the signal fall into the experimental region. In my examination, to date, the number of destination IP addresses is greater than 416,852. The number may have changed after June, 2004, since the data file I compiled seems to have increased dramatically after that period.

(5) *RTT*. This variable is to count how many round trip times (RTTs) can be attributed to each starting hop in each country by counting million-seconds for each destination. CAIDA has developed a special tool, Skitter, which actively probes forward IP paths and round trip times (RTTs) from a Skitter host to a specified list of destinations. They have deployed a number of monitors around the world. Each Skitter monitor continuously sends probe packets to destinations in its target list.

The number of times each destination is probed per day depends primarily on the total number of destinations in the target list and, to a lesser extent, on the current global conditions of the network. This is the most stable variable supplied by CAIDA.

Table 9: The descriptive table of "RTT" for 10 countries (1998–2007)

Year	US	UK	CN	JP	CA	KR	FR	SE	NL	NZ
1998	206.06	301.30	391.70	424.20	280.40	355.00	368.15	150.30	125.50	393.00
1999	173.27	277.98	336.80	348.24	220.28	345.19	269.20	140.60	121.00	356.00
2000	135.09	176.63	290.30	305.43	239.32	234.76	198.05	130.90	116.50	319.00
2001	124.83	141.79	252.20	170.46	126.57	184.92	149.60	86.97	94.08	282.00
2002	107.15	126.32	222.50	168.31	108.58	179.04	115.88	142.23	118.71	245.00
2003	105.65	103.21	201.20	160.97	103.52	148.58	80.28	123.36	119.35	193.02
2004	90.70	114.66	188.30	81.51	100.93	85.00	98.71	106.12	94.31	193.30
2005	66.96	78.67	106.27	43.70	66.67	40.00	82.28	75.44	83.70	119.73
2006	52.62	62.37	114.25	39.15	53.22	39.00	64.35	71.39	89.50	93.79
2007	45.39	53.24	102.39	30.66	41.63	35.00	62.09	59.97	85.00	60.00
Mean	110.77	143.62	220.59	177.26	134.11	164.65	148.86	108.73	104.77	225.48
Std Dev	51.42	85.49	99.11	139.28	83.19	119.77	101.33	33.22	16.76	113.24
Skewness	0.48	0.89	0.31	0.57	0.66	0.45	1.17	−0.19	−0.06	−0.01
Kurtosis	2.33	2.49	1.99	1.99	2.00	1.93	3.14	1.48	1.23	1.76

Last four rows are statistics for measurable variable "RTT" across time for each country during 1998–2007

.

(6) *HopDistance*. This counts how many hops before the signal successfully goes through from the source to many destinations. This is an alternative way to test the stability and the efficiency of information traveling. We know exactly where the destination IP address is located, and the number of hops before reaching the destination, which gives a sight to the different choices of tracking. Specially, if the signal sending fails, the HopDistance is zero, so we can calculate the ratio of failure too.

(7) *Reply TTL*. Skitter data also measures the forward IP paths to record each hop from a source to many destinations by incrementing the so-called "time to live"

(TTL) of each IP paths and recording replies from each router (or hop) leading to the destination hosts. This is also an alternative way to compare the connectivity of the Internet across times in different countries, compared to only counting how many web hosts are attributed to each country by counting top-level host domain names. Measurement of Reply TTL begins from June 2004, and with some testing traces being done in 2003, which count for some small numbers in the records, as Table 11 shows.

Table 10: The descriptive table of "HopDistance" for 10 countries (1998–2007)

Year	US	UK	CN	JP	CA	KR	FR	SE	NL	NZ
1998	12.19	11.53	12.13	12.69	12.51	11.74	9.94	10.88	9.71	15.46
1999	11.16	9.45	11.51	11.43	11.55	11.88	9, 88	10.76	9.61	14.92
2000	11.47	10.11	10.72	13.70	13.48	9.95	9.82	10.64	9.52	14.38
2001	9.90	10.50	9.89	8.23	11.88	14.22	9.76	7.94	8.54	13.84
2002	10.89	9.68	9.10	11.00	13.14	9.81	10.78	12.44	10.31	13.30
2003	11.52	10.45	8.46	10.83	10.50	10.67	7.76	11.54	10.10	12.22
2004	13.85	15.58	8.08	6.18	15.82	10.18	10.60	12.94	8.06	14.35
2005	13.45	12.47	9.61	10.07	11.29	9.92	11.39	10.19	9.19	11.58
2006	10.36	10.54	11.28	10.82	9.68	9.66	9.43	10.18	8.95	11.35
2007	9.02	9.60	9.73	10.05	7.22	9.40	9.20	8.80	8.85	10.60
Mean	11.38	10.99	10.05	10.50	11.71	10.74	9.86	10.63	9.28	13.20
Std Dev	1.50	1.86	1.34	2.12	2.33	1.48	0.99	1.51	0.70	1.66
Skewness	0.24	1.65	0.08	−0.59	−0.20	1.39	−0.57	−0.23	−0.20	−0.23
Kurtosis	2.29	4.77	1.87	3.05	3.06	3.94	3.38	2.47	2.14	1.67

Last four rows are statistics for measurable variable "HopDistance" across time for each country during 1998–2007

.

4.3.2 Computing Methodology

We develop the software using C/C++ and FORTRAN together with an MPI library for the parallel statistical analysis of the Internet trafficing data from 8 cities around the world over a period of 8 years. Each set of monthly data from each city is

Table 11: The descriptive table of "TTL" for 10 countries (2004–2007)

Year	US	UK	CN	JP	CA	KR	FR	SE	NL	NZ
2004	94.98	104.56	50.60	57.69	106.18	–	84.42	84.18	84.32	107.14
2005	70.99	71.03	47.41	40.85	69.11	–	72.24	60.99	76.10	71.62
2006	57.35	56.82	52.12	36.53	56.45	–	57.66	57.80	68.40	58.34
2007	48.57	47.67	44.61	28.67	44.88	–	52.54	48.17	60.20	34.00
Mean	67.97	70.02	48.69	40.93	69.15	–	66.72	62.78	72.25	67.77
Std Dev	20.23	24.95	3.35	12.26	26.59	–	14.46	15.27	10.34	30.52
Skewness	0.52	0.69	−0.23	0.58	0.70	–	0.27	0.71	0.00	0.29
Kurtosis	1.82	1.95	1.52	2.01	2.00	–	1.49	2.09	1.66	1.90

Last four rows are statistics for measurable variable "TTL" across time for each country during 2004–2007

.

about 64 MB in size and needs to be analyzed initially for 7 categories. This 7-category data analysis can be parallelized on multiprocessors. The data is downloaded from its provider (SDSC) and can be located on the system where the software is running. This will allow us to easily access the monthly data and develop statistical methods to analyze the data. We will also develop methods to measure the series correlation between cities in a panel format. A successful development of this computational software can be followed by an MRAC proposal for larger computer resource allocation to allow for more productive data analysis.

Using the city-level databases downloaded from the Skitter tools of CAIDA, we collect data from a total of twenty cities scattered around the world. There are ten cities in US: Urbana (IL), Bethesda (MD), College Park (MD), Aberdeen (MD), Moffett Field (CA), Palo Alto (CA), San Jose (CA), San Diego (CA), Eugene (OR), and Washington, DC. The non-US cities are: Cambridge (UK), London (UK), Shenyang (CN),Tokyo (JP), Ottawa (CA), Taejon (KR), Paris (FR), Stockholm (SE), Amsterdam (NL), Auckland (NZ). All the files are the daily records of each monitor from the round-trip tracing. We use the daily files at the end of each month to approximate the monthly behavior, because we have checked and found that, within each month, there is no big fluctuation across different daily file. Compiled files are transferred to be ASCII code for reading, and extracted to text files which only contain the details for the seven variables we are interested in. On the platform of PSC, especially the Fed, and Pople systems, we use parallel computing to deal with the data mining and numerical analysis. All of the text files for the twenty cities contain continuous behavior of signals, and the contents for each variable is in large volume. We average the instant measurements of each variable according to a specific number of points, to guarantee that each city will have the same amount of derived information for each variable in each month. Furthermore, the monthly data collected is translated into annual data for a 1998-2007 multivariate dynamic time series. (We use optimal interpolation tools for solving the missing data problem.) Finally, we use city-level databases to approximate the diffusion of the Internet in different countries. Those cities we collected from information are reasonably big within each country for approximation, ignoring the scale effect in the measurement.

4.3.3 Descriptive Results

After performing the above computing methodology, we find some patterns in the characteristics of these variables during 1998-2007. For the US in 1998, the size of the data packets sent out from the source monitor was 109.38 on average, and the standard deviation was 53.30, with the minimum and maximum size 26 and 174.52. In 2006, the size of the data packets increased to 174.52, and the standard deviation tripled. The minimum and maximum size of packets tripled too. This trend remains throughout 2007. This means that the capacity for sending information packets has been improved and the fluctuation also has more room to carry out.

Examining the the number of RTTs in 1998 and 2007, results correspond to our estimation: the Round Trip Time of one information message sent from the source monitor to different destination routers gets decreasing on average. The usage time drops from 206.06 to 45.39 milliseconds. This means that, in 2007, we could send more information per second and, in one second, we could get more information transferred to different countries in the world.

Table 12: The descriptive table of "Bilateral Trade Volume" for 9 countries (1998–2007)

Year	US	UK	CN	JP	CA	KR	FR	SE	NL	NZ
1998	–	73, 896.40	85, 409.80	179, 676.10	329, 859.50	40, 427.10	41, 744.80	11, 670.20	26, 577.00	3, 531.40
1999	–	77, 644.50	94, 899.30	188, 329.80	365, 311.10	54, 136.80	44, 586.20	12, 353.00	27, 911.40	3, 671.90
2000	–	84, 915.70	116, 203.40	211, 403.80	409, 779.20	68, 137.70	50, 161.60	14, 150.70	31, 506.70	4, 050.40
2001	–	82, 082.90	121, 460.70	183, 924.60	379, 692.00	57, 362.20	50, 272.50	12, 449.80	29, 000.00	4, 309.70
2002	–	73, 949.60	147, 320.30	172, 877.80	370, 010.40	58, 147.40	47, 256.00	12, 369.40	28, 159.10	4, 094.80
2003	–	76, 622.90	180, 804.00	170, 040.90	391, 518.40	61, 302.00	46, 272.30	14, 342.60	31, 647.80	4, 250.80
2004	–	82, 175.50	231, 109.80	183, 373.90	446, 239.70	72, 354.60	52, 523.40	15, 927.20	36, 590.20	5, 040.80
2005	–	89, 600.70	284, 662.10	192, 684.30	502, 283.00	71, 353.00	56, 100.70	17, 536.40	41, 329.70	5, 747.30
2006	–	98, 923.10	341, 447.40	206, 639.80	533, 093.90	78, 022.70	60, 551.40	17, 996.30	48, 301.90	5, 922.60
2007	–	106, 839.00	384, 379.80	206, 622.90	565, 944.90	81, 964.00	68, 228.70	17, 496.60	51, 240.10	5, 831.00
Mean	–	84, 665.03	198, 769.66	189, 557.39	429, 373.21	64, 320.75	51, 769.76	14, 629.22	35, 226.39	4, 645.07
Std Dev	–	10, 939.30	106, 788.80	14, 523.15	79, 460.78	12, 462.64	7, 992.12	2, 444.96	8, 883.17	914.19
Skewness	–	1.12	0.71	0.32	0.67	−0.45	0.97	0.27	0.96	0.43
Kurtosis	–	0.49	−0.94	−1.24	−0.90	0.00	0.68	−1.76	−0.51	−1.58

Values are millions. Last four rows are statistics for "Bilateral Trade Volume" across time for each country during 1998–2007

.

With respect to the number of hops to a specific destination router, we refer to the variable, HopDistance. The results are promising too, since the average number of hops drops from 12.19 to 9.02. This means that the traveling paths on the road towards the destination get less, and the time for information sent to the destination decreased as well. We get the support from the better stability and the efficiency of the information traveling, which get the same direction for the inference as that of RTT.

Finally, for the forward IP paths to record each hop form a source to many destination, we look at the variable, Reply TTL. The results are exciting, since the mean of the path number drops from 94.98 to 48.57. This means that the traveling paths on the road towards the destination get less indeed. This is a better way to compare the connectivity of the Internet across time in different country, compared to only count how many web hosts are attributed to each country by counting top-level host domain names.

Table 13: The descriptive table of "Real GDP" for 10 countries (1998–2007)

Year	US	UK	CN	JP	CA	KR	FR	SE	NL	NZ
1998	10,245.60	1,854.05	1,248.36	3,908.33	958.33	570.52	1,855.76	294.29	538.17	77.78
1999	10,701.44	1,910.08	1,343.25	3,907.57	1,011.35	624.64	1,915.47	307.77	559.67	81.81
2000	11,093.21	1,982.86	1,456.12	4,018.27	1,064.58	677.65	1,992.94	321.08	579.08	83.68
2001	11,176.49	2,029.50	1,577.01	4,024.69	1,085.02	703.65	2,028.23	324.45	587.34	86.59
2002	11,355.14	2,071.30	1,720.49	4,035.15	1,120.61	752.70	2,050.31	330.86	590.68	90.60
2003	11,640.13	2,126.51	1,892.53	4,094.16	1,139.77	776.01	2,072.25	335.71	585.49	93.86
2004	12,063.88	2,195.89	2,083.64	4,205.41	1,174.95	812.71	2,119.00	347.78	593.91	97.99
2005	12,433.39	2,238.32	2,300.31	4,284.16	1,207.10	846.83	2,155.43	357.17	603.00	99.90
2006	12,790.92	2,300.15	2,567.15	4,378.84	1,240.56	889.11	2,202.76	373.07	621.12	102.17
2007	13,064.85	2,359.02	2,859.80	4,467.88	1,274.22	933.32	2,252.54	383.14	643.56	105.43
Mean	11,656.51	2,106.77	1,904.87	4,132.45	1,127.65	758.71	2,064.47	337.53	590.20	91.98
Std Dev	917.82	167.27	540.40	193.78	100.54	116.19	123.91	28.05	29.39	9.38
Skewness	0.14	0.00	0.56	0.58	−0.21	−0.12	−0.19	0.22	0.07	−0.07
Kurtosis	−0.97	−1.09	−0.79	−0.87	−0.72	−0.85	−0.53	−0.65	0.71	−1.36

Values are billions. Last four rows are statistics for "Real GDP" across time for each country during 1998–2007

.

In summary, there are descriptive results with this Internet dataset:

1. Conducted on a high-speed testbed at the San Diego Supercomputer Center, the Skitter data available on www. caida.org falls into our field to experiment the Internet distance on the multiple outcomes.
2. We have about 20 cities from different countries scattered around the world, and

each of them send out the fixed-length message to about 1 million destination hosts, either located in developed countries, or developing countries.

Table 14: The descriptive table of "Populations" for 10 countries (1998–2007)

Year	US	UK	CN	JP	CA	KR	FR	SE	NL	NZ
1998	281.083	58.522	1245.990	126.286	30.124	45.755	58.610	8.847	15.735	3.798
1999	284.529	58.703	1256.730	126.500	30.398	46.110	58.847	8.848	15.826	3.830
2000	287.842	58.907	1266.950	126.706	30.687	46.429	59.128	8.860	15.915	3.868
2001	290.995	59.138	1276.680	126.907	30.993	46.707	59.459	8.886	16.001	3.912
2002	294.009	59.392	1285.980	127.097	31.315	46.948	59.832	8.924	16.084	3.962
2003	296.928	59.667	1294.940	127.263	31.646	47.164	60.230	8.970	16.164	4.013
2004	299.821	59.958	1303.670	127.384	31.979	47.366	60.630	9.018	16.241	4.064
2005	302.741	60.261	1312.250	127.449	32.307	47.566	61.013	9.066	16.316	4.111
2006	305.697	60.575	1320.720	127.451	32.628	47.766	61.373	9.113	16.389	4.153
2007	308.674	60.899	1329.090	127.396	32.945	47.962	61.714	9.159	16.460	4.193
Mean	295.23	59.60	1289.30	127.04	31.50	46.98	60.08	8.97	16.11	3.99
Std Dev	9.19	0.81	27.77	0.42	0.96	0.73	1.08	0.11	0.24	0.14
Skewness	−0.08	0.26	−0.13	−0.74	0.07	−0.34	0.13	0.48	−0.13	0.06
Kurtosis	−1.13	−1.19	−1.14	−0.88	−1.27	−0.91	−1.38	−1.27	−1.17	−1.42

Values are millions. Last four rows are statistics for "populations" across time for each country during 1998–2007

.

3. In Figure 20–21, data-length of sending message, RTT, Hopdistance, and TTL. We can see the overall pattern: for data-length, the message sending out to multiple destination gets increasing, RTT, Hopdistance, TTL gets downward sloping, indicating that the rounding time for the message echoing back to resource gets decreasing, so getting more efficient during the time range. This new measurement is more accurate compared to Freund(2004)[50].
4. In the natural experiment framework, we might control all other variables in the gravity equation, then concentrate on four variables for experiment group before and after June 2004. They said the number of destination and the distribution of destination hosts gets huge changes There might be more increment in hosts lo-

cated in developing countries, than in developed countries, since it's reasonable to see that he huge jump in hardware upgrade at end-points in developing countries, than developed countries. I will ask them about the details of those destination hosts.

Skitter data supply the packets form each monitor across through the different destinations, which can capture the effective volumes of information across different cities. This is a very useful measurement for us to test the relationship of the Internet effect on the multiple outcomes across different countries. To measure the information volumes of the Internet, CAIDA use the Skitter data to count how many packets (Pkts) are accumulated to each starting hop in each country by counting Pkts, Pkts/sec, Bytes, and Bits/sec for each destination.

It is shown that, based on Table 8-11, every country gets the same pattern as that of San Diego, as we illustrated above. We can also examine the discrepancy between the countries, due to the difference in the attributes of the Internet measurement. Especially, the economic globalization, social globalization, political globalization get the significant cross-sectional dependence. On the other hand, the change in testing basis for information through different source monitors around June, 2004 and persisting influence on all the variables can be also noticed. The bilateral trade volume, real GDP and population for ten countries during 1998-2007 are collected from international financial statistics supported by IMF.

4.4 MODEL AND RESULTS

4.4.1 Model

As long as we think about how the Internet, international trade and capital flow, social and political integration interact each other, we want to use the descriptive statistics to see the behaviors of different variables, and further to carry out the modeling issue, to build up a framework for the following estimation and testing issues.

Anderson (1979)[7] theoretically found the gravity model based on a "constant" elasticity of substitution preferences and differentiated goods based on the place of the origin of manufacturing. They assumed that the production or the supply is fixed for each good, so the theoretical background begins with the consumers' utility function, which is approximated by a CES utility function

$$\left(\sum_i \beta_i^{\frac{1-\sigma}{\sigma}} c_{ij}^{-\frac{1-\sigma}{\sigma}}\right)^{-\frac{\sigma}{1-\sigma}},$$

subject to the budget constraint

$$\sum_i p_{ij} c_{ij} = y_j.$$

where σ is the elasticity of substitution between all goods. β_i is a positive distribution parameter, c_{ij} is consumption by consumers in region j, of goods from region i, y_j is the nominal income of consumers in region j, p_{ij} is the price of goods which consumed by j consumers and are from region i. $p_{ij} = p_i t_{ij}$, where t_{ij} is the trade cost factor.

The nominal demand derived from maximization condition is

$$d_{ij} = \left(\frac{\beta_i p_i t_{ij}}{p_j}\right)^{1-\sigma} y_j,$$

where

$$p_j = \left[\sum_i \left(\beta_i p_i t_{ij}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}},$$

so based on market clearance,

$$\begin{aligned} y_j &= \sum_i x_{ij} = \left(\sum_i \frac{\beta_i p_i t_{ij}}{p_j}\right)^{1-\sigma} y_j \\ &= \left(\beta_i p_i\right)^{1-\sigma} \sum_i \left(\frac{t_{ij}}{p_j}\right)^{1-\sigma} y_j, \qquad \forall i. \end{aligned}$$

the derived gravity equation,

$$x_{ij} = \frac{y_i y_j}{\sum_j y_j} y_j \left(\frac{t_{ij}}{p_i p_j} \right)^{1-\sigma}$$

which implies that, for bilateral trade, x_{ij}, trader barriers reduce size-adjusted trade between large countries more than between small countries, and trader barriers raised size-adjusted trade within small countries more than within large countries.

Now, the most interesting part is to find the factors of the Internet to affect the trade volume. We applied the following,

$$t_{ij} = b_{ij} d_{ij} I_i I_j.$$

where $b_{ij} = 1$ if $i = j$ and is dropped in dynamic panel data because this issue has been taken care of in econometric setup, d_{ij} is the geographic distance, and $I_i I_j$ is the Internet measurement between region i and j. In this paper, we legitimately concentrate on the influence of the Internet on the international trade, especially the bilateral trade volume.

Now the fundamental empirical equation is

$$\ln x_{ij} = \text{constant} + (1-\sigma)\rho \ln d_{ij} + (1-\sigma) \ln I_i I_j + \ln y_i y_j - (1-\sigma) \ln pop_i pop_j,$$

where ρ is kind of correlation coefficient, and the GDP is treated as real GDP per capita by denoting the product of the populations of country i and j as $pop_i pop_j$, then it corresponds to

$$X_{ijt} = \beta_0 + \beta_1 D_{ijt} + \beta_3 I_i I_j + \beta_4 Y_{it} Y_{jt} + \beta_5 POP_{it} POP_{jt} + \beta_8 K_{ijt} + v_{ijt}. \tag{4.4.1}$$

This is the equation which has the same fundamentals as those in Freund and Weinhold (2000). X_{ijt}, for $t = 1998, ..., 2007$, is log of the bilateral trade volume between country i and j, or the international trade either including or excluding non-

Internet trading flow, or the degree of cross–country integrations, which are mainly derived from the international trade and capital flows, which can also be used as the sub-categories for the dependent variables in the above equation. I_iI_j is log of the product of the Internet distance; $Y_{it}Y_{jt}$ is log of the product of the real GDP of country i and j; POP_iPOP_j is log of the product of the populations of country i and j; K_{ijt} is the other control variables, mainly including generic divergence, and the dummy variable for big-country effect[10]; v_{ijt} is the error term of estimation equation. Here, we put all the variables which are not our interests into the constant, which, in term, can be treated as the intercept of the estimation equations.

4.4.2 Results

According to the equation (4.4.1), we apply the Arellano-Bond methods for panel data to make the estimation for the relationship between the bilateral trade volume and the diffusion of the Internet. Consider the equation (4.4.1) again, D_{ijt} was dropped because the geographic distance remains the same across countries which can be treated as fixed constant in the estimation equation. With the data described in section 4.3[11]:

$$\Delta X_{ijt} = \delta_0 + \delta_1 \Delta X_{ij,t-1} + \delta_2 \Delta I_{it}I_{jt} + \delta_3 \Delta Y_{it}Y_{jt} + \delta_4 \Delta POP_{it}POP_{jt} + u_{ijt}, \quad (4.4.2)$$

where the subscript i, j and t denote the ith and jth country, $i, j = 1, ..., 9$, and the tth year, $t = 1, ..., 10$. ΔX_{ijt} is the first-differenced bilateral trade volume for country i, in year t, along with the lagged terms of X_{ijt}. $\Delta I_{it}I_{jt}$ is the first–differenced Internet distance. In this paper, they are the measurements for the bandwidth estimations of the cross-traffic traveling distance, RTT or other closed-form measurements, for

[10] The dummy variables, K_{ijt}, are summarized into the intercept if we are not in the position to discuss these concerns.

[11] As previous literatures, we firstly run the level–valued estimations based on the above equation 4.4.2, and test the Auto–Regressive error terms for AR(1), AR(2). Our results show that there exists large significant error correlations. We set up the empirical bases for estimation with the equation 4.4.2.

instance, Data_Length, HopDistance, and TTL. Similar illustrations are for $\Delta Y_{it} Y_{jt}$ and $\Delta POP_{it} POP_{jt}$. u_{ijt}, the error term is assumed as multi-Gaussian, and orthogonal to the group of variables for identification.

Table 15: The results by using "Bilateral Trade Volume" as dependent variable

	(1)	(2)	(3)	(4)
Lag-1 Trade Volume	0.1769	0.2966*	0.4325*	0.4625*
	(0.0928)	(0.0934)	(0.0958)	(0.1125)
Data length$_i$ $*$ Data length$_j$ at time t	0.0386*	–	–	–
	(0.0081)	–	–	–
RTT$_i$ $*$ RTT$_j$ at time t	–	−0.0124*	–	–
	–	(0.0054)	–	–
HopDistance $_i$ $*$ HopDistance $_j$ at time t	–	–	0.1051*	–
	–	–	(0.0314)	–
TTL$_i$ $*$ TTL$_j$at time t	–	–	–	0.0589*
	–	–	–	(0.0107)
GDP$_i$ $*$ GDP$_j$ at time t	1.2588*	1.1815*	1.0306*	0.7427*
	(0.1868)	(0.2012)	(0.1980)	(0.2129)
POP$_i$ $*$ POP$_j$ at time t	0.8390	1.1632	−2.4689*	−1.8734
	(0.8647)	(0.9840)	(1.2553)	(1.1202)
AB AR(1)	−0.6500	−0.6700	−1.1500	−0.1300
	[0.5175]	[0.5009]	[0.2484]	[0.8953]
AB AR(2)	−1.6100	−1.3800	−0.6200	−0.3200
	[0.1065]	[0.1671]	[0.5338]	[0.7498]

Notes: (1) Table is set for the results from the model according to the equation (4.6.1), applying Arellano-Bond methods for cross-sectional time series to make the estimation for the relationship between the bilateral trade volume and the diffusion of the Internet, here, Data_Length, RTT, HopDistance, TTL. The first column contains the independent variables we examined. Intercepts are suppressed. All the entries are the values of estimators correspond to the independent variable in each row. Results here for comparison are GMM estimates. (2) Arellano-Bond test that average autocovariance in residuals of order 1 and 2 is almost 0 failed in 5% level for all alternative specifications. "$*$" represents the 5% significance. The paraphrased figures are the simple standard deviations, and the bracketed ones are the p–values.

Firstly, $X_{ij,t-1}$ and $I_{it} I_{jt}$ are treated as exogenous. We assume that $I_{it} I_{jt}$, $i = 1, ..., N$, for all $t = 1, 2, ...T$ can be assumed to be exogenous variables, and the additional orthogonality restrictions can be obtained according to strict or weak exogeneity of $X_{ij,t-1}$ and $I_{it} I_{jt}$.

There has evidence supporting the relationship between the bilateral trade volume and the diffusion of the Internet. We can find the strong dependence within the current and lagged values of bilateral trade volume, which is significantly shown up in the estimation equation. Here, the coefficient estimates of $X_{ij,t-1}$ are positive in all the alternative specifications.Table (15) shows that the coefficients are 0.1769, 0.2966, 0.4325, and 0.4625 for including independent variable, Data_length, RTT, HopDistance, TTL, respectively, besides the population variables. They are significant for 5% level. The positive and consistent coefficient estimates of $X_{ij,t-1}$ confirm that the current year's measurable increment for bilateral trade volume is adaptive to lagged-one-year's values. Moreover, the insignificant coefficient estimates of $X_{ij,t-2}$ reject the lagged-two-year's dependence and are supportive of the model in Table (15). The signs of those estimates are insignificant, which are omitted for reporting. So in this setup framework, the relationship we examined for series $X_{ij,t-1}$ are statistically significant in most cases. Now the most important part is to examine the impact from these new Internet measurements. Data_length, RTT, HopDistance and TTL are all significant. We can find that, the Internet measurement, Data_length, has positive relationship in estimation process. The current change in the diffusion of the Internet has statistically significant relationship to the current bilateral trade volume. Bilateral trade volume is increased by 0.386 % associated with ten percentage increasing in the measures of size of information transformation. Note that in our setup, all the other measurements, RTT, HopDistance and TTL are strongly significant in all model alternatives. 0.124 % increase in bilateral trade volume measurements are associated with ten percentage decreasing in the measures of data-packets-transferred speed time within one round-trip-time. We can recall that RTT is the most stable variable supplied from the CAIDA in my previous data description section. 1.051 % increase in bilateral trade volume is associated with 10 % increasing in the measures of numbers of hops transferred during the information transfer, and 0.589 % increase in bilateral trade volume is associated with 10 % increasing in the measures of numbers of forward IP paths transferred within one round-trip-time. These results confirms the usefulness of this unique data sets introduced by this paper, and also effectively suggests a new way for measuring the diffusion of the Internet and the impact on the level of the bilateral trade volume across border [12].

On the other hand, in the examination of real GDP, we have positive coefficients supporting the relationship between the bilateral trade volume and the diffusion of the Internet. Here, the coefficient estimates of $Y_{it}Y_{jt}$ are positive in all the alternative specifications, in the Table (15). The results show that the coefficients are 1.2588, 1.1815, 1.0306, and 0.7427. They are significant for 5% level. The positive and consistent coefficient estimates of $Y_{it}Y_{jt}$ confirm that the current year's measurable increment for bilateral trade volume is adaptive to current growth of GDP across countries. The magnitude of coefficients associated with GDP is larger than one except the last column. It's expected that the increased growth rate of bilateral trade volume is larger that the increased growth rate of GDP. We can find that, the current changes in growth rate of the population have no statistically significant relationship to the current changes in growth rate of bilateral trade volume, The third column is the only exception. The coefficient, -2.4689, is statistically significant in the estimation process and the sign is negative, which exactly matches our need to include $POP_{it}POP_{jt}$ in estimation equation to capture the "per capita" effect. The absolute value of -2.4689 is larger than the absolute value of 1.0306. We can expect that the positive impact from the growth of GDP is diluted by the growth of population, and the dilution is much greater than the positive effect from GDP.

Based on our results, aggregate cross-sectional time series data shows consistent evidence to provide support for the dependence between the bilateral trade volume and the Internet diffusion. Arellano-Bond test that average autocovariance in residuals of order 2 is 0 fail in 5% level for all columns, so putting the lagged Internet measurement series into the explainary sets have been statistically approved. The probability of not rejecting the hypothesis of average autocovariance in residuals of order 2 is 0 range from 0.11 to 0.75, which is explicitly showing that the higher-order

[12] Freund and Weinhold (2000) concluded that the Internet coefficient in 1999 suggests that a ten percent increase in the number of hosts in one country would have led to about 0.3–0.9 percent greater trade in different specifications.

residuals yield no autocorrelation. *The impact of the Internet diffusion on bilateral trade volume is supported in the empirical results by using this unique data set.* we may argue that it will become reasonable if we use the Internet diffusion rate to calculate the network incentive, not the proxy variable which results from the global geography, local distance and neighboring price for services.

4.5 ELASTICITY ANALYSIS AND MOMENT CONDITION SPECIFICATION

We might be interested in measuring how the change in the diffusion of the Internet affects the quantity of bilateral trade volume. attempting to develop such summary measures is that bilateral trade volume and the diffusion of the Internet are not measured in the same units. The diffusion of the Internet is measured in second per information package anuually, and the bilateral trade volume is measured in billion dollars. We might then speak of a fall of 10 % in the measures of data-packets-transferred speed time within one round-trip-time, leading an increase of 0.124 % in bilateral trade volume. Similarly, we could speak of an increase in the measures of size of information transformation, leading to an increase in bilateral trade volume of 0.386 % growth as was the case in RTT case. However ,there now would be no easy way to answer the question of whether bilateral trade volume is more or less responsive to RTT changes than to Data_length changes.

After addressing the detailed results of model specification, I discuss here the elasticity of growth rate of bilateral trade volume and the growth rate of the diffusion of the Internet. Table (16) is set for the elasticity results from the model according to the coefficients in the equation (4.6.1). Applying elasticity formula, $\delta_2 \Delta I_{it} I_{jt} / \Delta X_{ijt}$, for cross-sectional time series to make the estimation for the relationship between the bilateral trade volume and the diffusion of the Internet, we make some discussion

Table 16: The Elasticity of Bilateral Trade Volume versus Internet Measurements

Elasticity	Data length	RTT	HopDistance	TTL
	0.0466	−0.0144	0.2412	0.0783

Notes: (1) Table is set for the elasticity results from the model according to the coefficients in the equation (4.6.1). Applying elasticity formula, $\delta_2 \Delta I_{it} I_{jt} / \Delta X_{ijt}$, for cross-sectional time series to make the estimation for the relationship between the bilateral trade volume and the diffusion of the Internet, here, Data_Length, RTT, HopDistance, TTL.

here for elasticity of growth rate of bilateral trade volume and the growth rate of the diffusion of the Internet. Table (16) show that they are 0.0466, 0.0144, 0.2412, and 0.0783 for Data_length, RTT, HopDistance and TTL, respectively. They are all significantly bigger than the magnitude of coefficients associated with each Internet measurement. It's shown that the scaled percentage changes in bilateral trade volume is even greater than the scaled percentage changes of the diffusion of the Internet in the period 1998–2007. This shows exactly how the bilateral trade volume responds, ceteris paribus, to a 10 percent change in each Internet measurement. This gives us another insight for the impact of the Internet on the international trade. Although the partial derivative, δ_2, also shows how bilateral trade volume changes when the Internet measurement changes, it is not as useful as the elasticity because it is measured in units of bilateral trade volume per unit change in the Internet measurement. In the elasticity, multiplication of that partial derivative by $\Delta I_{it} I_{jt} / \Delta X_{ijt}$ causes the units to "drop out", and the remaining expression is purely in terms of proportions. In our cases here, we might know that a 10 percent change in the size of information transferred leads to a 0.466 percent change in the growth of bilateral trade volume, whereas a 10 percent change in the measures of data-packets-transferred speed time within one round-trip-time, leads to a 0.144 percent change in the growth of bilateral trade volume. Consequently, we could conclude that growth of bilateral trade volume were more responsive to the size of information transferred. Furthermore, 2.412 %

increase in growth of bilateral trade volume is associated with a 10 % changes in the measures of numbers of hops transferred during the information transfer, and 0.783 % increase in growth of bilateral trade volume is associated with 10 % changes in the measures of numbers of forward IP paths transferred within one round-trip-time.

On the other hand, the potential endogeneity problem inferring from that gravity equations illustrated above needs to be discussed too. I assume that there is no third factor that causes the Internet traffic to increase and later causes the bilateral trade measure to increase. In the absence of such a cause, then I state that one causes the other. Actually, we firstly run the level–valued estimations based on the above equation 4.4.2, and test the Auto–Regressive error terms for AR(1), AR(2). Our results show that there exists large significant error correlations. Arellano-Bond tests show that average autocovariance in residuals of order 2 is statistically 0. The probability of not rejecting the hypothesis of average autocovariance in residuals of order 2 is 0 range from 0.11 to 0.75, which is explicitly showing that the higher-order residuals yield no autocorrelation.

When the problem of omitted variables comes out, it is reasonable to lead to bias which are caused by the endogeneity issues. Dragging lagged-one-period bilateral trade volume inside the structural equation can avoid the omitted variable problem, and make the Internet measurements to be valid variables to estimate the bilateral international digital trade in the time period. Let's reinvestigate the descriptive results of this data set, which are illustrated before in section 4.3.1. The measurements for each country in period 1998–2007 are listed in Table 8–11, where the last four rows are statistics of annual data covering the time period. The identification issue is one of the basic step for estimation, if we want to use the data to capture the correlations between the variables interested. In our empirical model above, we assume that the error terms which cannot be observed are multi-Gaussian distribution conditional on the Internet variables. The linear correlations in the regressions count heavily on the multi–dimensional distribution of errors conditional on the explanatory variables, where the regressions can begin towards the estimations and inferences. The (weak) orthogonal conditions between the error terms and the explanatory groups are the bases for consistency and efficiency issues, which is, of course, the reason for the potential bias in the maximum likelihood estimation process. We can anticipate that the relationship between the Internet measurement and the international trade can be treated as multivariate Gaussian.

4.6 DYNAMIC PANEL CAUSALITY AND SENSITIVITY ANALYSIS

To study the direction of dependency between two random variables, a very powerful tool is the Granger Causality test, but this method is checking the lagged terms of pairwise multivariate time series in vector autoregression. In the sequel, I want to turn the attention to the dynamic panel causality through quasi-sensitivity analysis of the variables $\{X_{ijt}\}$ and $\{I_{it}I_{jt}\}$. Arellano-Bond methods for cross-sectional time series make the estimation for the causality of the diffusion of the Internet, here, Data_Length, RTT, HopDistance, TTL, on the bilateral trade volume, which allow us to separate the effect of the dependence from the inverse effects of the independent variables.

According to the equation (4.4.2), we apply the Arellano-Bond methods for panel data to make the estimation for the relationship between the diffusion of the Internet and the bilateral trade volume. Consider the equation (4.4.2) again inversely, with the data described in section 4.3:

$$\Delta I_{it}I_{jt} = \delta_0+\delta_1\Delta I_{i,t-1}I_{j,t-1}+\delta_2\Delta X_{ijt}+\delta_3\Delta X_{ij,t-1}+\delta_4\Delta Y_{it}Y_{jt}+\delta_5\Delta POP_{it}POP_{jt}+\epsilon_{ijt}, \tag{4.6.1}$$

where the subscript i, j and t denote the ith and jth country, $i, j = 1, ..., 9$, and the tth year, $t = 1, ..., 10$. ΔX_{ijt} , $\Delta I_{it}I_{jt}$, $\Delta Y_{it}Y_{jt}$ and $\Delta POP_{it}POP_{jt}$ are previously denoted. $\Delta I_{i,t-1}I_{j,t-1}$ is the lagged-one-period terms of $\Delta I_{it}I_{jt}$; ϵ_{ijt}, the error term is assumed as multi-Gaussian, and orthogonal to the group of variables for identification.

Table 17: The results by using "Bilateral Trade Volume" as independent variable

	Data Length	RTT	HopDistance	TTL
Lag-1 Data length$_i$ $*$ Data length$_j$	0.5191*	–	–	–
	(0.1556)	–	–	–
Lag-1 RTT$_i$ $*$ RTT$_j$	–	0.6781*	–	–
	–	(0.1130)	–	–
Lag-1 HopDistance$_i$ $*$ HopDistance$_j$	–	–	0.1925	–
	–	–	(0.1357)	–
Lag-1 TTL$_i$ $*$ TTL$_j$	–	–	–	0.0185
	–	–	–	(0.2461)
Trade Volume at time t	1.6919	−0.2420	0.3237	0.6257
	(0.8918)	(0.4247)	(0.3936)	(0.5786)
Lag-1 Trade Volume	−1.3970	−0.3816	−1.4122*	0.9746
	(0.8101)	(0.3555)	(0.3433)	(0.9471)
GDP$_i$ $*$ GDP$_j$ at time t	1.0176	0.4855	1.5486	0.1608
	(1.4959)	(0.8928)	(0.7881)	(1.5037)
POP$_i$ $*$ POP$_j$ at time t	−1.5435	4.6670	0.1428	−1.5164
	(7.3670)	(3.7767)	(3.5250)	(6.9736)
AB AR(1)	−4.9400	−2.4700	−4.2300	0.2700
	[0.0000]	[0.0156]	[0.0000]	[0.7900]
AB AR(2)	2.0500	−1.9500	0.3400	–
	[0.0408]	[0.0511]	[0.7352]	–

Notes: (1) Table is set for the results from the model according to the equation (4.6.1), applying Arellano-Bond methods for cross-sectional time series make the estimation for the causality of the diffusion of the Internet, here, Data_Length, RTT, HopDistance, TTL, on the bilateral trade volume. The first column contains the independent variables we examined. Intercepts are suppressed. All the entries are the values of estimators correspond to the independent variable in each row. Results here for comparison are GMM estimates. (2) Arellano-Bond test that average autocovariance in residuals of order 1 is almost 0 not failed in 5% level for all alternative specifications. Arellano-Bond test that average autocovariance in residuals of order 2 is almost 0 failed in 5% level for all alternative specifications. "*" represents the 5% significance. The paraphrased figures are the simple standard deviations, and the bracketed ones are the p–values.

There exists lagged-one-period causality from the Internet on the current values of related variables, but Arellano-Bond test that average autocovariance in residuals of order 1 is almost 0 not failed in 5% level for all alternative specifications. Here, the coefficient estimates of $I_{i,t-1}I_{j,t-1}$ are positive in first two of the alternative specifications. Table (17) shows that the coefficients are 0.5191, 0.6781, 0.1925, and 0.0185 for including other dependent variable, bilateral trade volume, besides the lagged values, GDP and population for both country i and j, respectively. All the other coefficient estimates confirm that the reverse relationship for the Internet measurement and international trade are not significant at all.

Based on our results, dynamic panel causality and sensitivity analysis demonstrate further empirically that the causality of the Internet diffusion on the bilateral trade volume in (4.4.2). Arellano-Bond test that average autocovariance in residuals of order 1 is 0 didn't fail in 5% level for almost all columns, so the estimation equation has serial autocovariance in error terms. The probability of not rejecting the hypothesis of average autocovariance in residuals of order 2 is 0 is also high, which is showing that the higher-order residuals yield some autocorrelation. In Table (17), it should be noticed that both the $X_{ij,t}$, the current level of bilateral trade volume and the $X_{ij,t-1}$, the lagged values, have no relationship with the current Internet diffusion, which support the impact from the Internet diffusion on the global integration, not the inverse way. we may argue here again that how this unique new data set of the Internet measurement can improve the calculation for growth of bilateral trade volume across countries.

4.7 CONCLUSION

With the measurement of the Internet cross-traffic traveling distance, we evaluate the effect of the Internet distance on international trade, especially on the bilateral trade volume with multiple outputs. Under the development of the Internet, network construction, and information technology, the evaluation of modern international trade gets new supportive documents.

The major finding is that there exists significant and positive relationship between the Internet distance and the international trade. Comparing cross-country technology reform based on the examination from the Internet service, we find the persistent impact on the bilateral trade. The GDP matters continuously across countries, but there is no significant effect from the population to the bilateral trade volume.

Elasticity analysis further supports the importance of the diffusion of the Internet. Dynamic panel causality and sensitivity analysis demonstrate further empirically that the causality of the Internet diffusion on the bilateral trade volume empirically in (4.4.1). The causality issue points out more research views, such as how this unique new data set of the Internet measurement can improve the calculation for the relationship between the Internet and the bilateral trade volume in a panel way. On the other hand, as long as we know the level of bilateral trade volume for developed countries and developing countries, this unique data set is an efficient alternative for estimating the relationship between the trading pattern and the diffusion of the Internet, which will be beneficial to examine export and import opportunities in the global cooperation.

5.0 ACKNOWLEDGEMENTS

The author would like to thank participants at the thesis defense. Special Thanks are to Xinfu Chen, my adviser, for intelligent directing throughout the paper, and Bruce Qiang Sun, my husband, for supporting all the run.

6.0 APPENDIX

A.1 PROOF OF THEOREM 1

Let $t \geqslant 0$ be an arbitrary time. We wait till the current time is t and then measure values $\mu_t^T, \sigma_{t1}^T, \cdots, \sigma_{tn}^T$ for all $T \geqslant t$. Set $\mathbf{a}^T = (1, \sigma_{t1}^T, \cdots, \sigma_{tn}^T)$ and let m be the dimension of the set $\{\mathbf{a}^T \mid T > t\} \subset \mathbb{R}^{n+1}$. Then $1 \leqslant m \leqslant n+1$ and there exist $T_1 > t, \cdots, T_m > t$ such that $\mathbf{a}^{T_1}, \cdots, \mathbf{a}^{T_m}$ are linearly independent. Let $T_{m+1} > t$ be arbitrarily. We can find $(v_1, \cdots, v_{m+1}) \subset \mathbb{R}^{m+1}$ such that

$$\sum_{i=1}^{m+1} v_i \mathbf{a}^{T_i} = \mathbf{0}, \qquad \sum_{i=1}^{m} |v_i| > 0, \qquad \sum_{i=1}^{m+1} \mu_t^{T_i} v_i \geqslant 0. \tag{A.1.1}$$

Now consider an investment of buying $v_i / Z_t^{T_i}$ share of T_i-bond for $i = 1, \cdots, m+1$ at time t and sell all of them at time $t+dt$. The total initial cost of such an investment is

$$\sum_{i=1}^{m+1} \frac{v_i}{Z_t^{T_i}} \, Z_t^{T_i} = \sum_{i=1}^{m+1} v_i = 0,$$

by the first component of the equation $\sum_{i=1}^{m+1} v_i \mathbf{a}^{T_i} = \mathbf{0}$. After selling all these bonds, we obtain at time $t + dt$ a profit of the amount

$$P_{t+dt} = \sum_{i=1}^{m+1} \frac{v_i}{Z_t^{T_i}} \, Z_{t+dt}^{T_i} = \sum_{i=1}^{m+1} v_i \frac{Z_{t+dt}^{T_i}}{Z_t^{T_i}} - \sum_{i=1}^{m+1} v_i \frac{Z_t^{T_i}}{Z_t^{T_i}} = \sum_{i=1}^{m+1} v_i \frac{dZ_t^{T_i}}{Z_t^{T_i}}.$$

Using (2.2.1) we find that

$$P_{t+dt} \;=\; \sum_{i=1}^{m+1} v_i \Big\{ \mu_t^{T_i}\,dt + \sum_{k=1}^{n} \sigma_{t\,k}^{T_i}\,dX_t^k \Big\} = \Big(\sum_{i=1}^{m+1} v_i \mu_t^{T_i}\Big) dt + \sum_{k=1}^{n} \Big(\sum_{i=1}^{m+1} v_i \sigma_{t\,k}^{T_i}\Big)\, dX_t^k.$$

Since the equation $\sum_{i=1}^{m+1} v_i \mathbf{a}^{T_i} = \mathbf{0}$ implies $\sum_{i=1}^{m+1} v_i \sigma_{t\,k}^{T_i} = 0$ for every $k = 1, \cdots, n$, we, starting from nothing, obtained a profit of $P_{t+dt} = (\sum_{i=1}^{m+1} v_i \mu_t^{T_i})dt \geqslant 0$ at time $t + dt > t$. By the no-arbitrage assumption, there must hold $\sum_{i=1}^{m+1} v_i \mu_t^{T_i} = 0$. In view of (A.1.1) and the definition of $\mathbf{a}^T$, we see that

$$m \geqslant \operatorname{rank} \begin{pmatrix} 1 & \cdots & 1 & 1 \\ \sigma_{t\,1}^{T_1} & \cdots & \sigma_{t\,1}^{T_m} & \sigma_t^{T_{m+1}} \\ \vdots & \cdots & \vdots & \vdots \\ \sigma_{t\,n}^{T_1} & \cdots & \sigma_{t\,n}^{T_m} & \sigma_{tn}^{T_{m+1}} \\ \mu_t^{T_1} & \cdots & \mu_t^{T_m} & \mu_t^{T_{m+1}} \end{pmatrix} \geqslant \operatorname{rank} \begin{pmatrix} 1 & \cdots & 1 \\ \sigma_{t\,1}^{T_1} & \cdots & \sigma_{t\,1}^{T_m} \\ \vdots & \cdots & \vdots \\ \sigma_{t\,n}^{T_1} & \cdots & \sigma_{t\,n}^{T_m} \end{pmatrix} = m \tag{A.1.2}$$

By rearranging indexes $\{1, \cdots, n\}$ if necessary we can assume that the first m rows in the second matrix above is linearly independent. Then we find unique constants $C^1, \cdots, C^m$ that depend only on $t, T_1, \cdots, T_m$ such that

$$(\mu_t^{T_1}, \cdots, \mu_t^{T_m}) = (1, \cdots, 1)C^1 + \sum_{k=1}^{m-1} (\sigma_{t\,k}^{T_1}, \cdots, \sigma_{t\,k}^{T_m}) C^{k+1}.$$

The first inequality in (A.1.2) then implies that $\mu_t^{T_{m+1}} = C^1 + \sum_{k=1}^{m-1} \sigma_{t\,k}^{T_{m+1}} C^{k+1}$. Since $C^1, \cdots, C^m$ do not depend on T_{m+1}, setting $T = T_{m+1}$, $R_t = C^1$, $P_t^i = C^{i+1}$ for $i = 1, \cdots, m-1$ and $P_t^i = 0$ for $i \geqslant m$ we obtain (2.2.2). This completes the proof.

A.2 PROOF OF THEOREM 2

We divide the proof in several steps.

1. Set $\sigma_t^{ij} = \mathrm{Cov}(dX_t^i, dX_t^j)/dt$. Differentiate (2.2.3) by Itô's Lemma [77] to obtain

$$\frac{dZ_t^T}{Z_t^T} = \Big\{A_0'(s) + \sum_{i=1}^n A_i'(s)X_t^i + \frac{1}{2}\sum_{i,j=1}^n \sigma_t^{ij} A_i(s)A_j(s)\Big\}dt - \sum_{i=1}^n A_i(s)dX_t^i$$

where $s = T - t$ and $A_i'(s) = dA_i(s)/ds$. This implies that Z_t^T satisfies (2.2.1) with

$$\sigma_{ti}^T = -A_i(s), \quad \mu_t^T = A_0'(s) + \sum_{i=1}^n A_i'(s)X_t^i + \frac{1}{2}\sum_{i,j=1}^n \sigma_t^{ij} A_i(s)A_j(s)\ .$$

Hence, by TSM, there are $\{R_t, P_t^1, \cdots, P_t^n\}$ such that

$$A_0'(s) + \sum_{i=1}^n A_i'(s)X_t^i + \frac{1}{2}\sum_{i,j=1}^n \sigma_t^{ij} A_i(s)A_j(s) = R_t - \sum_{i=1}^n P_t^i A_i(s). \tag{A.2.1}$$

For convenience, in the sequel, we set $X_t^0 \equiv 1$ for all t. Also, $\tau \geqslant 0$ is a fixed time such that the covariance matrix $\mathrm{Cov}(X_\tau^1, \cdots, X_\tau^n)$ is positive definite.

2. Since $Z_\tau^\tau = 1$, we obtain from (2.2.3) that $0 = A_0(0) + \sum_{i=1}^n A_i(0)X_\tau^i$. This implies that

$$A_i(0) = 0 \quad \forall\, i = 0, \cdots, n,$$

since $\mathrm{Cov}(X_\tau^1, \cdots, X_\tau^n)$ is positive definite. Consequently, setting $s = 0$ in (A.2.1) we obtain

$$R_t = \sum_{k=0}^n r_k X_t^k \qquad \forall\, t \geqslant 0, \qquad r_k = A_k'(0) \quad \forall\, k = 0, \cdots, n.$$

3. Using a linear regression we write (defining $X_t^0 \equiv 1$)

$$P_\tau^i = \sum_{k=0}^n p_k^i X_\tau^k + \varepsilon_\tau^i, \qquad \sigma_\tau^{ij} \;=\; \sum_{k=0}^n \sigma_k^{ij} X_\tau^k + \eta_\tau^{ij}$$

where p_k^i, σ_k^{ij} are constants (that may depend on τ) and ε_τ^i and η_τ^{ij} are random variables satisfying $\mathbb{E}[\varepsilon_\tau^i X_\tau^k] = 0 = \mathbb{E}[\eta_\tau^{ij} X_\tau^k]$ for all $i, j = 1, \cdots, n$ and $k = 0, \cdots, n$. Multiplying (A.2.1) evaluated at $t = \tau$ by X_τ^i and taking the expectation we then obtain

$$\sum_{k=0}^{n} \Big\{ A_k'(s) - r_k + \sum_{i=1}^{n} p_k^i A_i(s) + \frac{1}{2} \sum_{i,j=1}^{n} \sigma_k^{ij} A_i(s) A_j(s) \Big\} \mathbb{E}[X_\tau^k X_\tau^i] = 0 \quad \forall\, i = 0, \cdots, n.$$

Since $\mathrm{Cov}[X_\tau^1, \cdots, X_\tau^n]$ is positive definite, we then obtain

$$A_k'(s) = r_k - \sum_{i=1}^{n} p_k^i A_i(s) - \frac{1}{2} \sum_{i,j=1}^{n} \sigma_k^{ij} A_i(s) A_j(s) \quad \forall\, k = 0, \cdots, n. \tag{A.2.2}$$

4. For each t, define

$$\varepsilon_t^i := P_t^i - \sum_{k=0}^{n} p_k^i X_t^k, \qquad \eta_t^{ij} := \sigma_t^{ij} - \sum_{k=0}^{n} \sigma_k^{ij} X_t^k, \quad i, j = 1, \cdots, n.$$

Substituting (A.2.2) into (A.2.1) we obtain

$$\frac{1}{2} \sum_{i,j=1}^{n} \eta_t^{ij} A_i(s) A_j(s) + \sum_{i=1}^{n} \varepsilon_t^i A_i(s) = 0 \quad \forall\, s \geqslant 0.$$

Now we assume that the $(n+1)(1+n/2)$ functions $A_k(s), A_i(s)A_j(s), k = 0, \cdots, n, i = 1, \cdots, n, j = 1, \cdots, i$, are linearly independent. Since $\eta_t^{ij} = \eta_t^{ji}$, we see that $\varepsilon_t^i = \eta_t^{ij} \equiv 0$ for all $i, j = 1, \cdots, n$. This completes the proof.

A.3 PROOF OF THEOREM 4

Set $F_t = (F_t^1, \cdots, F_t^n)$, $\sigma_{tk}^T = -L_k(T-t)$, $P_t^k = P^k(F_t)$ and $R_t = R(F_t)$.

Let $s^1, \cdots, s^n$ be positive numbers such that the matrix $\mathbf{L}(s^1, \cdots, s^n)$ in (2.2.8) is non-singular. By continuity, there is a positive constant h such that $\mathbf{L}(s^1+\delta, \cdots, s^n+\delta)$ is also non-singular for every $\delta \in [0, h]$. Set $T^i = T + s^i$ for $i = 1, \cdots, n$.

Let $t_0 \in [T-h, T)$ be an arbitrary fixed time. Consider in time interval $[t_0, T]$ a dynamic portfolio that stars with cash $V(F_{t_0}, t_0)$ and for each trading period $(t, t+dt)$ holds $w_{ti}/Z_t^{T_i}$ shares of T^i-bond for $i = 1, \cdots, n$ and the rest money in the $(t+dt)$-bond. The portfolio is managed up to time T by the repeated process of buying the required shares of bonds at time t, selling it at time $t+dt$ and immediately buying again according to the new required shares. A **trading strategy** is a prescription of observable weights. Here we consider a strategy where the weights $w_{t1}, \cdots, w_{tn}$ are the solutions of the linear system

$$\sum_{k=1}^{n} w_i \sigma_{tk}^{T^i} = \frac{\partial V(z,t)}{\partial z^k}\Big|_{z=F_t} \quad \forall\, k = 1, \cdots, n. \tag{A.3.1}$$

Since the matrix $(\sigma_{tk}^{T^i})_{n\times n} = \mathbf{L}(s^1+\delta, \cdots, s^n+\delta)$ with $\delta = T - t$ is non-singular, there is a unique solution. Also as the right-hand side is observable at time t, so is the weight $(w_{t1}, \cdots, w_{tn})$. Hence, the strategy is executable, i.e., a trading strategy.

Now we calculate the value, denoted by V_t, of the portfolio at any time $t \in [t_0, T]$. At time t there are $w_{ti}/Z_t^{T_i}$ shares of T^i-bond and $[V_t - \sum_{i=1}^n w_{ti}]/Z_t^{t+dt}$ shares of $t+dt$-bond, so the change of the value of the portfolio from t to $t+dt$ is

$$\begin{aligned} dV_t &= V_{t+dt} - V_t = \sum_{i=1}^{n} \frac{\omega_{ti}}{Z_t^{T^i}} Z_{t+dt}^{T^i} + \frac{V_t - \sum_{i=1}^n \omega_{ti}}{Z_t^{t+dt}} Z_{t+dt}^{t+dt} - V_t \\ &= \sum_{i=1}^{n} w_{ti} \frac{dZ_t^{T^i}}{Z_t^{T^i}} + \Big(V_t - \sum_{i=1}^{n} w_{ti}\Big) \frac{Z_{t+dt}^{t+dt} - Z_t^{t+dt}}{Z_t^{t+dt}} \\ &= \sum_{i=1}^{n} w_{ti} \Big\{ R_t dt + \sum_{k=1}^{n} \sigma_{tk}^{T^i} \Big(P_t^k dt + dF_t^k\Big)\Big\} + \Big(V_t - \sum_{i=1}^{n} w_{ti}\Big) R_t dt \end{aligned}$$

by (2.2.7). Using (A.3.1) and the fact that $R_t = R(F_t), P_t^k = P^k(F_t)$ we derive that

$$\begin{aligned} dV_t &= \Big\{V_t R_t + \sum_{k=1}^{n} P_t^k \Big(\sum_{i=1}^{n} w_{ti} \sigma_{tk}^{T_i}\Big)\Big\} dt + \sum_{k=1}^{n} \Big(\sum_{i=1}^{n} w_{ti} \sigma_{ti}^{T_i}\Big) dF_t^k \\ &= V_t R_t dt + \sum_{k=1}^{n} P^k(z) \frac{\partial V(z,t)}{\partial z^k}\, dt + \sum_{k=1}^{n} \frac{\partial V(z,t)}{\partial z^k} dF_t^k \Big|_{z=F_t}. \end{aligned}$$

Using (2.2.11) to replace the second term and writing $V(z,t)$ as V we obtain

$$dV_t = V_t R_t dt + \Big(\frac{\partial V}{\partial t} + \frac{1}{2}\sum_{i,j=1}^{n} \sigma^{ij}(z)\frac{\partial^2 V}{\partial z^i \partial z^j} - R(z)V\Big)dt + \sum_{k=1}^{n} \frac{\partial V}{\partial z^k} dF_t^k\Big|_{z=F_t}.$$

Since $\sigma^{ij}(F_t) = \mathrm{Cov}(dF_t^i, dF_t^j)/dt$ and V is assumed to satisfy the Itô Lemma, we have

$$dV(F_t,t) = \sum_{k=1}^{n} \frac{\partial V(z,t)}{\partial z^k} dF_t^k + \Big(\frac{\partial V(z,t)}{\partial t} + \frac{1}{2}\sum_{i,j=1}^{n} \sigma^{ij}(z)\frac{\partial^2 V(z,t)}{\partial z^i \partial z^j}\Big)dt\Big|_{z=F_t}.$$

Hence,

$$dV_t = [V_t - V(F_t,t)]R_t dt + dV(F_t,t)).$$

An integration gives $[V_t - V(F_t,t)] = [V_{t_0} - V(F_{t_0},t_0)]e^{\int_{t_0}^t R_s ds} = 0$ for all $t \in [t_0,T]$. Thus, at time T, the portfolio worths

$$V_T = V(F_T,T) = \Phi(F_T) = P_T.$$

That is, the value of the portfolio equals exactly the payment of the security derivative. Such a portfolio is called a **replication portfolio**. Since the replication portfolio pays exactly the security derivative at time T, by no arbitrage, the price of the portfolio at time t_0 is $V_{t_0} = V(F_{t_0},t_0)$. As $t_0 \in [T-h,T)$ is arbitrary, the value of the security derivative is $V(F_t,t)$ for any $t \in [T-h,T]$. Hence, the assertion of theorem holds for any $t \in [T-h,T)$.

Now consider $t \in [T-2h, T-h)$. The security derivative can be regarded as a time $T-h$ payment of $P_{T-h} = \Phi^h(z)|_{z=F_{T-h}}$ where $\Phi^h(\cdot) := V(\cdot, T-h)$. Hence, applying the assertion just established, we see that the value of the security derivative at time t is $V(F_t,t)$ for any $t \in [T-2h, T-h)$. Repeating the same argument we then obtained the assertion of the Theorem. This completes the proof.

A.4 PROOF OF THEOREM 6

1. Let $\{(\lambda_k, \mathbf{e}_k)\}_{k=1}^n$ be a complete eigenset of $\mathbf{C} = (\mathbb{E}[\xi^i \xi^j])_{m\times m}$ where $\{\lambda_k\}_{k=1}^m$ is in decreasing order and $\{\mathbf{e}_k\}_{k=1}^m$ is an orthonormal set. For $k = 1, \cdots, K :=$ $\dim\{\{\xi^1, \cdots, \xi^m\})$, we define $g^k := \sum_{i=1}^m \xi^i e_k^i / \sqrt{\lambda_k} \in (\{\xi^1, \cdots, \xi^m\}$. Then for $k, l =$ $1, \cdots, K$,

$$\langle g^k, g^l \rangle = \frac{1}{\sqrt{\lambda_k \lambda_l}} \sum_{j=1}^m \sum_{i=1}^m e_k^i \langle \xi^i, \xi^j \rangle e_l^j = \frac{\mathbf{e}_k \mathbf{C} \mathbf{e}_l'}{\sqrt{\lambda_k \lambda_l}} = \frac{\lambda_k \mathbf{e}_k \cdot \mathbf{e}_l}{\sqrt{\lambda_k \lambda_l}} = \delta^{kl}.$$

Thus, $\{g^1, \cdots, g^K\}$ is an orthonormal set and $(\{\xi^1, \cdots, \xi^m\}) = (\{g^1, \cdots, g^K\})$. In addition, $\langle \xi^i, g^k \rangle = \sum_{j=1}^m \mathbf{e}_k^j \langle \xi^j, \xi^i \rangle / \sqrt{\lambda_k} = (\mathbf{e}_k \mathbf{C})^i / \sqrt{\lambda_k} = \sqrt{\lambda_k} e_k^i$.

2. Let $n \in \{1, \cdots, K\}$ and V be an n-dimensional subspace of $(\{\xi^1, \cdots, \xi^m\})$. Let $\{f^1, \cdots, f^n\}$ be an orthogonal base of V and $\{f^1, \cdots, f^K\}$ be an orthonormal base of $\{\xi^1, \cdots, \xi^m\}$. Then $\xi^i = \sum_{k=1}^K \langle \xi^i, f^k \rangle f^k$ and

$$\sum_{i=1}^m \mathrm{dist}^2(\xi^i, V) \;=\; \sum_{i=1}^m \Big\| \sum_{k=n+1}^K \langle \xi^i, f^k \rangle f^k \Big\|^2 = \sum_{i=1}^m \sum_{k=n+1}^K \langle \xi^i, f^k \rangle^2 .$$

Substituting $\xi^i = \sum_{l=1}^K \langle \xi^i, g^l \rangle g^l = \sum_{l=1}^K \sqrt{\lambda_l}\, e_l^i\, g^l$ we obtain

$$\sum_{i=1}^m \mathrm{dist}^2(\xi^i, V) \;=\; \sum_{i=1}^m \sum_{k=n+1}^K \sum_{l=1}^K \sum_{l'=1}^K \sqrt{\lambda_l}\, e_l^i \langle g^l, f^k \rangle \langle f^k, g^{l'} \rangle e_{l'}^i \sqrt{\lambda_{l'}} = \sum_{k=n+1}^K \sum_{l=1}^K \lambda_l \langle f^k, g^l \rangle^2$$

since $\sum_{i=1}^m e_l^i e_{l'}^i = \mathbf{e}_l \cdot \mathbf{e}_{l'} = \delta_{ll'}$. By $1 = \|g^l\|^2 = \sum_{k=1}^K \langle g^l, f^k \rangle^2$ and $1 = \sum_{l=1}^K \langle g^l, F_k \rangle^2$,

$$\sum_{k=n+1}^K \sum_{l=1}^n \langle g^l, f^k \rangle^2 = \sum_{k=n+1}^K \Big[1 - \sum_{l=n+1}^K \langle g^l, f^k \rangle^2\Big] = \sum_{l=n+1}^K \Big[1 - \sum_{k=n+1}^K \langle g^l, f^k \rangle^2\Big] = \sum_{l=n+1}^K \sum_{k=1}^n \langle g^l, f^k \rangle^2.$$

Thus,

$$\begin{aligned}\sum_{i=1}^m \mathrm{dist}^2(\xi^i, V) \;&=\; \sum_{l=1}^K \lambda_l \sum_{k=n+1}^K \langle g^l, f^k \rangle^2 = \sum_{l=n+1}^K \lambda_l \Big[1 - \sum_{k=1}^n \langle g^l, f^k \rangle^2\Big] + \sum_{l=1}^n \lambda_l \sum_{k=n+1}^K \langle g^l, f^k \rangle^2 \\ &=\; \sum_{l=n+1}^K \lambda_l + \sum_{l=n+1}^K (\lambda_n - \lambda_l) \sum_{k=1}^n \langle g^l, f^k \rangle^2 + \sum_{l=1}^n (\lambda_l - \lambda_n) \sum_{k=n+1}^K \langle g^l, f^k \rangle^2.\end{aligned}$$

Thus, using $\lambda_n - \lambda_l \geqslant 0$ for $l \geqslant n$ and $\lambda_l - \lambda_n \geqslant 0$ for $l \leqslant n$, we have

$$\sum_{i=1}^{m} \mathrm{dist}^2(\xi^i, V) \geqslant \sum_{l=n+1}^{K} \lambda_l = \sum_{i=1}^{m} \mathrm{dist}^2(\xi^i, (\{g^1, \cdots, g^n\})).$$

Since V is an arbitrary n-dimensional subspace of $\{\xi^1, \cdots, \xi^m\})$, we see that

$$\min_{\dim(V)=n} \sum_{i=1}^{n} \mathrm{dist}^2(\xi^i, V) = \min_{\dim(V)=n, V\subset\{\xi^1,\cdots,\xi^m\}} \mathrm{dist}^2(\xi, V) = \sum_{l=n+1}^{K} \lambda_l.$$

Hence, $(\{g^1, \cdots, g^n\})$ is a principal subspace of $\xi^1, \cdots, \xi^m$. Consequently, $\{g^1, \cdots, g^K\}$ is a set of principal components of $\{\xi^1, \cdots, \xi^m\}$.

3. Finally, suppose $\{f^1, \cdots, f^K\}$ is a set of principal components. Set $V = (\{f^1, \cdots, f^n\})$. Then $\sum_{i=1}^{n} \mathrm{dist}^2(\xi^i, V) = \sum_{k=n+1}^{K} \lambda_l$ so that

$$\sum_{l=n+1}^{K} (\lambda_n - \lambda_l) \sum_{k=1}^{n} \langle g^l, f^k \rangle^2 + \sum_{l=1}^{n} (\lambda_l - \lambda_n) \sum_{k=n+1}^{K} \langle g^l, f^k \rangle^2 = 0.$$

This equation is true for every $n = 1, \cdots, K$. We can derive that $\langle g^l, f^k \rangle = 0$ if $\lambda_k \neq \lambda_l$. Thus, $f^k = \sum_{\lambda_l=\lambda_k} \langle f^k, g^l \rangle g^l = (\xi^1, \cdots, \xi^n)\tilde{\mathbf{e}}_k^\top / \sqrt{\lambda_k}$ where $\tilde{\mathbf{e}}_k$ is an eigenvector of $\mathbf{C}$ associated with λ_k. In addition, from $\delta^{kl} = \langle f^k, f^l \rangle = \tilde{\mathbf{e}}_k \mathbf{C} \tilde{\mathbf{e}}_l^\top / \sqrt{\lambda_k \lambda_l} = \mathbf{e}_k \mathbf{e}_l^\top \sqrt{\lambda_k / \lambda_l}$, we see that $\{\tilde{\mathbf{e}}_1, \cdots, \tilde{\mathbf{e}}_K\}$ is an orthonormal set. This completes the proof.

A.5 DERIVATION OF FORWARD EXCHANGE RATE DYNAMICS

The determination of exchange rates and currency returns is set up in the context of the linear term structure model (LTSM), and under covered interest parity: *The* *price B_t^T of the T-bond at time t satisfies*

$$f_t^T - s_t = r_t^T - r_t^{T*} := \frac{1}{T-t}\log\frac{1}{B_t^T} - \frac{1}{T-t}\log\frac{1}{B_t^{T*}}$$

$$= \frac{1}{T-t}\sum_{i=1}^{n} L_i(T-t)\, F_t^i - \frac{1}{T-t}\sum_{i=1}^{n} L_{i*}(T-t)\, F_t^i \qquad \forall\, t \geqslant 0, T \in [t, t+T^{\max}) \tag{A.5.1}$$

where $L_1(\cdot), \cdots, L_n(\cdot)$ are differentiable functions defined on $[0, T^{\max})$ and $\{(F_t^1, \cdots, F^n)\}$ is an Itô process with a positive definite matrix $(\mathbb{E}[F_\tau^i F_\tau^j])_{n\times n}$ for some $\tau \geqslant 0$ for both currencies, along with (3.3.9) and (3.3.10).

A.6 DERIVATION OF EXPECTED FUTURE SPOT RATE EXAMINATION AND RISK PREMIUM SETUP

By no–arbitrage condition,

$$1 \times e^{r_t^T(T-t)} = [\frac{S_T^{\mathbb{Q}}}{S_t}]e^{r_t^{T*}(T-t)}$$
$$e^{(r_t^T - r_t^{T*})(T-t)} = \frac{S_T^{\mathbb{Q}}}{S_t}. \tag{A.6.1}$$

here, suppose $\mathbb{Q}$ is a measure under which $\{\{s_T^{\mathbb{Q}}\}_{T\geq t}$ are martingales, then

$$\ln S_T^{\mathbb{Q}} = \ln S_t + (T-t)(r_t^T - r_t^{T*})$$
$$\ln S_T^{\mathbb{Q}} - \ln S_t = (T-t)\frac{[\sum_{i=1}^n L_i(T-t) - \sum_{i=1}^n L_i^*(T-t)]\, F_t^i}{T-t}$$
$$\ln S_T^{\mathbb{Q}} - \ln S_t = \sum_{i=1}^{n}[L_i(T-t) - L_i^*(T-t)]\, F_t^i. \tag{A.6.2}$$

where $\{(F_t^1, \cdots, F_t^n)\}$ is assumed to be a martingale under a measure $\mathbb{P}$ of physical observation, now, $S_T^{\mathbb{Q}}$ can be written as

$$S_T^{\mathbb{Q}} := S_T + \int_t^T K_\zeta d\zeta \qquad \forall\, T \geq t \geqslant 0. \tag{A.6.3}$$

Suppose $\mathbb{Q}$ is a measure under which $\{\{S_T^{\mathbb{Q}}\}\}_{T\geq t}$ are martingales. This particular measure $\mathbb{Q}$ is called the *risk-neutral measure*. The K_ζ is named as K-risk.

Now we can write future spot rate as

$$\begin{aligned}
\ln(S_T + \int_t^T K_\zeta^i d\zeta) &= \ln S_t + \sum_{i=1}^{n} [L_i(T-t) - L_i^*(T-t)]\ F_t^i \\
\ln(S_T + \int_t^T K_\zeta^i d\zeta) - \ln S_t &= \sum_{i=1}^{n} [L_i(T-t) - L_i^*(T-t)]\ F_t^i.
\end{aligned} \tag{A.6.4}$$

Similarly, in affine term structure, **ATSM**, the group of expected future spot exchange rate can be expressed as

$$\begin{aligned}
\ln S_T^{\mathbb{Q}} &= \ln S_t + (T-t)(r_t^{T*} - r_t^T), \\
\ln S_T^{\mathbb{Q}} - \ln S_t &= (T-t)[\frac{A_0 - A_0^*}{T-t} \\
&\quad + \frac{(\sum_{i=1}^n B_i(T-t) - \sum_{i=1}^n B_i^*(T-t))\ X_t^i}{T-t}] \\
\ln S_T^{\mathbb{Q}} - \ln S_t &= [A_0(T-t) - A_0^*(T-t)] \\
&\quad + \sum_{i=1}^{n} [B_i(T-t) - B_i^*(T-t)]\ X_t^i.
\end{aligned}$$

Now we can write future spot rate as

$$\begin{aligned}
\ln(S_T + \int_t^T K_\zeta^i d\zeta) &= \ln S_t + [A_0(T-t) - A_0^*(T-t)] + \sum_{i=1}^{n} [L_i(T-t) - L_i^*(T-t)]\ X_t^i \\
\ln(S_T + \int_t^T K_\zeta^i d\zeta) - \ln S_t &= [A_0(T-t) - A_0^*(T-t)] + \sum_{i=1}^{n} [L_i(T-t) - L_i^*(T-t)]\ X_t^i
\end{aligned} \tag{A.6.5}$$

which can be used to match the results of the previous literature.

On the other hand, we have forward premium as follows,

$$f_t - s_t = (T-t)(r_t^{T*} - r_t^T)$$
$$f_t - s_t = (T-t)\frac{\left[\sum_{i=1}^n L_i(T-t) - \sum_{i=1}^n L_i^*(T-t)\right] F_t^i}{T-t}$$
$$f_t - s_t = \sum_{i=1}^n [L_i(T-t) - L_i^*(T-t)]\, F_t^i. \tag{A.6.6}$$

so using term structure framework, we can define the forward premium as the factors multiplied by the loading difference between domestic and foreign currencies.

A.7 PROOF OF PROPOSITION 1

Following Chen and Huang (2008) [24]: **1.** Let $\{(\lambda_k, \mathbf{e}_k)\}_{k=1}^n$ be a complete eigenset of $\mathbf{C} = (\mathbb{E}[\xi^i\xi^j])_{m\times m}$ where $\{\lambda_k\}_{k=1}^m$ is in decreasing order and $\{\mathbf{e}_k\}_{k=1}^m$ is an orthonormal set. For $k = 1, \cdots, K := \dim\{\{\xi^1, \cdots, \xi^m\}\}$, we define $g^k := \sum_{i=1}^m \xi^i e_k^i/\sqrt{\lambda_k} \in (\{\xi^1, \cdots, \xi^m\}$. Then for $k, l = 1, \cdots, K$,

$$\langle g^k, g^l\rangle = \frac{1}{\sqrt{\lambda_k\lambda_l}}\sum_{j=1}^m\sum_{i=1}^m e_k^i\langle\xi^i,\xi^j\rangle e_l^j = \frac{\mathbf{e}_k\mathbf{C}\mathbf{e}_l'}{\sqrt{\lambda_k\lambda_l}} = \frac{\lambda_k\mathbf{e}_k\cdot\mathbf{e}_l}{\sqrt{\lambda_k\lambda_l}} = \delta^{kl}.$$

Thus, $\{g^1, \cdots, g^K\}$ is an orthonormal set and $(\{\xi^1, \cdots, \xi^m\}) = (\{g^1, \cdots, g^K\})$. In addition, $\langle\xi^i, g^k\rangle = \sum_{j=1}^m \mathbf{e}_k^j\langle\xi^j,\xi^i\rangle/\sqrt{\lambda_k} = (\mathbf{e}_k\mathbf{C})^i/\sqrt{\lambda_k} = \sqrt{\lambda_k}e_k^i$.

2. Let $n \in \{1, \cdots, K\}$ and V be an n-dimensional subspace of $(\{\xi^1, \cdots, \xi^m\})$. Let $\{f^1, \cdots, f^n\}$ be an orthogonal base of V and $\{f^1, \cdots, f^K\}$ be an orthonormal base of $\{\xi^1, \cdots, \xi^m\}$. Then $\xi^i = \sum_{k=1}^K\langle\xi^i, f^k\rangle f^k$ and

$$\sum_{i=1}^m \mathrm{dist}^2(\xi^i, V) = \sum_{i=1}^m\left\|\sum_{k=n+1}^K\langle\xi^i, f^k\rangle f^k\right\|^2 = \sum_{i=1}^m\sum_{k=n+1}^K\langle\xi^i, f^k\rangle^2.$$

Substituting $\xi^i = \sum_{l=1}^K \langle \xi^i, g^l\rangle g^l = \sum_{l=1}^K \sqrt{\lambda_l}\, e_l^i\, g^l$ we obtain

$$\sum_{i=1}^m \mathrm{dist}^2(\xi^i, V) \;=\; \sum_{i=1}^m \sum_{k=n+1}^K \sum_{l=1}^K \sum_{l'=1}^K \sqrt{\lambda_l}\, e_l^i \langle g^l, f^k\rangle \langle f^k, g^{l'}\rangle e_{l'}^i \sqrt{\lambda_{l'}} = \sum_{k=n+1}^K \sum_{l=1}^K \lambda_l \langle f^k, g^l\rangle^2$$

since $\sum_{i=1}^m e_l^i e_{l'}^i = \mathbf{e}_l \cdot \mathbf{e}_{l'} = \delta_{ll'}$. By $1 = \|g^l\|^2 = \sum_{k=1}^K \langle g^l, f^k\rangle^2$ and $1 = \sum_{l=1}^K \langle g^l, F_k\rangle^2$,

$$\sum_{k=n+1}^K \sum_{l=1}^n \langle g^l, f^k\rangle^2 = \sum_{k=n+1}^K \Big[1 - \sum_{l=n+1}^K \langle g^l, f^k\rangle^2\Big] = \sum_{l=n+1}^K \Big[1 - \sum_{k=n+1}^K \langle g^l, f^k\rangle^2\Big] = \sum_{l=n+1}^K \sum_{k=1}^n \langle g^l, f^k\rangle^2.$$

Thus,

$$\begin{aligned}\sum_{i=1}^m \mathrm{dist}^2(\xi^i, V) \;&=\; \sum_{l=1}^K \lambda_l \sum_{k=n+1}^K \langle g^l, f^k\rangle^2 = \sum_{l=n+1}^K \lambda_l \Big[1 - \sum_{k=1}^n \langle g^l, f^k\rangle^2\Big] + \sum_{l=1}^n \lambda_l \sum_{k=n+1}^K \langle g^l, f^k\rangle^2 \\ &=\; \sum_{l=n+1}^K \lambda_l + \sum_{l=n+1}^K (\lambda_n - \lambda_l) \sum_{k=1}^n \langle g^l, f^k\rangle^2 + \sum_{l=1}^n (\lambda_l - \lambda_n) \sum_{k=n+1}^K \langle g^l, f^k\rangle^2.\end{aligned}$$

Thus, using $\lambda_n - \lambda_l \geqslant 0$ for $l \geqslant n$ and $\lambda_l - \lambda_n \geqslant 0$ for $l \leqslant n$, we have

$$\sum_{i=1}^m \mathrm{dist}^2(\xi^i, V) \geqslant \sum_{l=n+1}^K \lambda_l = \sum_{i=1}^m \mathrm{dist}^2(\xi^i, (\{g^1, \cdots, g^n\})).$$

Since V is an arbitrary n-dimensional subspace of $\{\xi^1, \cdots, \xi^m\})$, we see that

$$\min_{\dim(V)=n} \sum_{i=1}^n \mathrm{dist}^2(\xi^i, V) = \min_{\dim(V)=n, V\subset\{\xi^1,\cdots,\xi^m\}} \mathrm{dist}^2(\xi, V) = \sum_{l=n+1}^K \lambda_l.$$

Hence, $(\{g^1, \cdots, g^n\})$ is a principal subspace of $\xi^1, \cdots, \xi^m$. Consequently, $\{g^1, \cdots, g^K\}$ is a set of principal components of $\{\xi^1, \cdots, \xi^m\}$.

3. Finally, suppose $\{f^1, \cdots, f^K\}$ is a set of principal components. Set $V = (\{f^1, \cdots, f^n\})$. Then $\sum_{i=1}^n \mathrm{dist}^2(\xi^i, V) = \sum_{k=n+1}^K \lambda_l$ so that

$$\sum_{l=n+1}^K (\lambda_n - \lambda_l) \sum_{k=1}^n \langle g^l, f^k\rangle^2 + \sum_{l=1}^n (\lambda_l - \lambda_n) \sum_{k=n+1}^K \langle g^l, f^k\rangle^2 = 0.$$

This equation is true for every $n = 1, \cdots, K$. We can derive that $\langle g^l, f^k\rangle = 0$ if $\lambda_k \neq \lambda_l$. Thus, $f^k = \sum_{\lambda_l=\lambda_k} \langle f^k, g^l\rangle g^l = (\xi^1, \cdots, \xi^n)\tilde{\mathbf{e}}_k^\top / \sqrt{\lambda_k}$ where $\tilde{\mathbf{e}}_k$ is an eigenvector of $\mathbf{C}$

associated with λ_k. In addition, from $\delta^{kl} = \langle f^k, f^l \rangle = \tilde{\mathbf{e}}_k \mathbf{C} \tilde{\mathbf{e}}_l^\top / \sqrt{\lambda_k \lambda_l} = \mathbf{e}_k \mathbf{e}_l^\top \sqrt{\lambda_k / \lambda_l}$, we see that $\{\tilde{\mathbf{e}}_1, \cdots, \tilde{\mathbf{e}}_K\}$ is an orthonormal set. This completes the proof.

BIBLIOGRAPHY

[1] Abramson, B. D., "Current Statistics: Internet Globalization Indicators", *telecommunications Policy*, Vol. **24**, (2000), 69–74.

[2] Ahn, D.H., Dittmar, R.F., Gallant, A.R., "Quadratic term structure models: theory and evidence", *Review of Financial Studies* **15**, 243-288, 2001.

[3] Ahn, D.H., "Common Factors and Local Factors: implications for term structrues and Exchange Rates", *Journal of Financial and Quantitative Analysis* **39, 1**, 69–102, March, 2004.

[4] Alexander, C., *Market Models: A Guide to Financial Data Analysis*, Wiley, Hoboken, NJ. 2001.

[5] Alvarez, F., Atkeson, A., and Kehoe, P. J., "Time-Varing Risk, Interest Rates, and Exchange Rates in General Equilibrium", *Review of Economic Studies* **76**, 851–878, March, 2009.

[6] Amin, K., and Jarrow, R.A., "Pricing foreign currency options under stochastic interest rates", *Journal of International Money and Finance*, **10**, 310-329, 1991.

[7] Anderson, J. and Wincoop, E., "Gravity with Gravitas: A Solution to the Border Puzzie", *American Economic Review*, **93:1** (2003), 170-192.

[8] Andres, L., Cuberes, D., Diouf, M. A., and Serebrisky, T., "Diffusion of the Internet: A Cross-country Analysis", Policy Research Working Paper, **4420**, Dec. 2007.

[9] Antonelli and Honeyman (2000), "Wiretapping the Internet', SPIE's Symposium on Enabling Technologies.

[10] Babbs, S. H., and Nowman, B. K., "Kalman Filtering of Generalized Vasicek Term Structure Models", *Journal of Financial and Quantative Analysis*, **34(1)**, 115-129, March, 1999.

[11] Babbs, S. H., and Nowman, B. K., "Econometric Analysis of a Continuous Time Multi-Factor Generalized Vasicek Term Structure Model: International Evidence", *Asia-Pacific Financial Markets*, **5**,159-183, 1998.

[12] Backus, D., Foresi, S., Telmer, C., "Interpreting the forward premium anomaly", *Canadian Journal of Economics* **28**, S108-S119, 1995.

[13] Backus, D. K., Foresi, S., and Telmer, C. I., "Affine Term Structure Models and the Forward Premium Anomaly", *Journal of Finance*, **LVI, 1**, 279–304, Feb., 2001.

[14] Bakshi, G.S., and Chen, Z., "Equilibrium valuation of foreign exchange claims", *Journal of Finance* **52**, 799-822, 1997.

[15] Basal, R., "An exploration of the forward premium puzzle in the currency markets", *Review of Financial Studies* **10**, 369-403, 1997.

[16] Baum, C., and Bekdache, B., "Factor-GARCH Modeling of the Treasury Term Structure", working paper, Boston College, June, 1996.

[17] Bhandari, A. K., and Heshmati, A., "Measurement of globalizaiton and Its Variations Among Countries, Regions adn Over Time", IZA discussion paper series (DP) No. **1578**, April, 2005.

[18] Black, F., Derman, E., and Troy, W., "A One-Factor Model on Interest Rates and its Ajpplication to Treasury Bond Options", *Financial Analysis Journal*, **46**, 33-39, 1990.

[19] Black, F., Karasinski, P., "Bond and Option Pricing when Short Rates are Lognormal", *Financial Analysis Journal*, 52-59, July-August, 1991.

[20] Black, F., and Scholes, M., "The Pricing of Options and Corporate Liabilities", *The Journal of Political Economy*, **81, 3**, 637–654, 1973.

[21] Brooks, C., "Linear and Nonlinear (Non-)Forecastability of High-Frequency Exchange Rates", *Journal of Forecasting*, **16**,125–145, 1997.

[22] Campbell, J. Y., Lo, A. W., and MacKinlay, A. C., *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ, 1997.

[23] Chacko, G. and Das, S., "Pricing Interest Rate Derivatives: A General Approach", *The Review of Financial Studies*, **15, 1**, 195–241, 2002.

[24] Chen, X. and Huang, T. T., "Theoretical and Empirical Analysis of Linear Term Structure Models", Working Paper, August 2008.

[25] Choudhry, T. & Lawler, P., "The Monetary Model of the Exchange Rate: Evidence from the Canadian Float of the 1950's", *Journal of Macroeconomics*, **19, 2**,349–362, 1997.

[26] Christopher J. N. & Lucio S., "How Well Do Monetary Fundamentals Forecast Exchange Rates?", working paper, the Federal Reserve of St. Louis, 2002.

[27] Clarke, G. and Wallsten, S., "Has The Internet Increased Trade? Developed And Developing Country Evidence", *Economic Inquiry*, **44(3)**, July 2006, 465–484.

[28] Cochrane, J.H., *Asset Pricing*, Princeton University Press, Princeton, NJ,2001.

[29] Cochrane, J. H., and Piazzesi, M., , "Bond Risk Premia ", *American Economic Review*, **95, 1**, March 2005, 138–160.

[30] Connor, G. , "The Three Types of Factor Models: A comparison of Their explanatory Power", *Financial Analysts Journal*, **51**, 42-46, 1995.

[31] Cox, J. C., Ingersoll, J. E., and Ross, S. A., "A Theory of The Term Structure of Interest Rates", *Econometrica*, **53(2)**, 385-407, Mar., 1985.

[32] Dai, Q., Singleton, K. J., "Specification Analysis of Affine term Structure Models", *Journal of Finance*, **55**, 1943-1978, 2000.

[33] Dai, Q., Singleton, K. J., "Term Structure Dynamics in Theory and Reality", *The Review of Financial Studies*, **16, 3**, 631–678, Fall, 2003.

[34] Dai, Q., Singleton, K. J., "Discrete–Time Dynamic Term Structure Models with Generalized Market Prices of Risk", *Working Paper*,April, 2006.

[35] Daniels, J. P. & Vanhoose, D. D., Chapters 2,4,5, *International Monetary & Financial Economic*, 2nd ed. South Western 2002.

[36] Dewachter, H., and Maes, K., "An Affine Model for International Bond Markets", *Working Paper*, 2001.

[37] Diebold, F. X., Piazzesi, M., and Rudebusch, G. D. , "Modeling Bond Yields in Finance and Macroeconomics", *American Economic Review P & P*, **95, 2**, May 2005, 415–420.

[38] Dreher, A., "Does globalization Affect Growth?", working paper, January, 2003.

[39] Dreher, A., Gaston, N., and Martens, P., , *Measuring Globalisation*, Springer, June, 2008.

[40] Duffie, D., Filipovic, D., and Schachermayer, W., "Affine Processes and Applications in Finance", *The Annals of Applied Probability*, **13**, 984-1053, 2003.

[41] Duffie, D., Kan, R., "Multi-Factor Term Structure Models", Phil. Trans. R. Soc. Lond. A, **347**, 577-586, 1994.

[42] Duffie, D., Kan, R., "A Yield-Factor Model of Interest Rates", *Mathematical Finance*, **6**, 379-406, 1996.

[43] Duffie, D., Pan, J., Singleton, K. J., "Transform Analysis and Asset Pricing for Affine Jump-Diffusions", *Econometrica*, **6**, 1343-1376, Oct.,2000.

[44] Duffee, G. R., "Term Premia and Interest Rate Forecasts in Affine Models", *Journal of Finance*, **LVII, 1**, 405–443, Feb., 2002.

[45] Fama, E., "Forward and spot exchange rates", *Journal of Monetary Economics*, **14**, 319-338, 1984.

[46] Firoiu, Boudec, Towsley, and Zhang (2002), "Theories and Models for Internet Quality of Service", Proceedings of the IEEE, May 2002.

[47] Frankel, J. A., "Monetary and Portfolio–Balance Models of the Determinations of Exchange Rates", Chapter 4, *On Exchange Rate*, MIT Press, Cambridge, 1993.

[48] Frankel, J. A. & Rose, A. K., "A Panel Project On Purchasing Power Parity: Mean Reversion Within and Between Countries", *Journal of International Economics*, **40**,209–224, 1996.

[49] Freund, C. and Weinhold, D., "An Empirical Investigation of the Internet and International Trade: The Case of Bolivia", working paper, Dec. 2000.

[50] Freund and Weinhold, "The Effect Of The Internet On International Trade", FRB International Finance Discussion Paper, No. 693, Dec. 2000.

[51] Froot, K. & Rogoff, K., "Perspectives on PPP and Long-Run Real Exchange Rates", in Grossman, G. and Rogoff, K. (eds.), *Handbook of International Economics*, Vol.3, Amsterdam: North-Holland, 1647–1688, 1995.

[52] Gagiardini, P., and Gourieroux, C., “An Efficient Nonparametric Estimator For Models With Nonlinear Dependence”, *Journal of Econometrics*, **137**, 189–229, 2007.

[53] Gallant, A. R., & Tauchen, G., “Which Moments to Match?”, *Econometric Theory*, **12**, 657–681, 1996.

[54] Greenstein, S., and Prince, J., ”The Diffusion of the Internet and the geography of the Digital Divide in the United States”, NBER working paper series, April, 2006.

[55] Goldfarb, A. and Blum, B. S., ”Does the Internet Defy the Law of Gravity?”, *Journal of International Economics*, Vol. **70 (2)**, December, 2006, 384–405.

[56] Gourieroux, C., & Sufana, R., “A classification of Two–Factor Affine Diffusion Term Structure Models”, *Journal of Financial Econometrics*, **4, 1**, 31–52, 2006.

[57] Heath, D., Jarrow, R., & Morton, A., “Bond Pricing and the Term Structrue of Interest Rates: A New Methodology for Contingent Claims Valuation”, *Econometrica*, **60**, 77-106, 1992.

[58] Ho, T., & Lee, S., “Term Structure movements an pricing interest rate contingent claims”, *Journal of finance*, **42**, 1129-1142, 1986.

[59] Hodrick, R., Vassalou, M., “Do we need multi–country models to explain exchange rate and interest rateand bond return dynamics?” *Journal of Economic Dynamics and Control* **26**, 1275-1299, 2002.

[60] Hofstader, ”Internet Accessibility: Beyond Disability”, webtechnologies, Sep. 2004.

[61] Huang, T. T., “Theoretical and Empirical Analysis of Common Factors in a Term Structure Model”, MS thesis, Univeristy of Pittsburgh, 2008.

[62] Husted, S. & Macdonald, R., “ Monetary-based Models of the Exchange Rate: A Panel Approach”, *Journal of International Financial Markets, Institutions & Money*, 1–20, January 1998.

[63] Inoue, A., Killian, L., “On the selection of forecasting models”, *European Central Bank Working Paper Series*, **214**, 2003.

[64] Inci, A. C., and Lu, B., "Exchange Raes and Interesty Rates: Can Term Structure Models Explain currency Movements?", *Journal of Economic Dynamics & Control*, **28**(2004), 1595-1624.

[65] Jong, F. D., "Time Series and Cross-Section Information in Affine Term-structure Models", *Journal of Business & Economic Statistics*, **18(3)**, 300-314, July, 2000.

[66] Junker, M., Szimayer, A., and Wagner, N., "Nonlinear term structure dependence: Copula Functions, Empirics, and Risk Implications", *Journal of Banking & Finance*, **30**(2006), 1171-1199.

[67] Juselius, K., "Do Purchasing Power Parity and Uncovered Interest Parity Hold in the Long–Run? An Example of Likelihood Inference in a Multivariate Time–Series Model", *Journal of Econometrics*, **69**,211–240, 1995.

[68] Knez, P.J., Litterman, R., Scheinkman, J., "Explorations into factors explaining money market returns", *Journal of Finance*, **49**, 1861-1882, 1991.

[69] Kouretas, G. P., "Identifying Linear Restrictions on the Monetary Exchange Rate Model and the Uncovered Interest Parity: Cointergration evidence from the Canadian-U.S. Dollar", *Canadian Journal of Economics*, No.4a, 1997.

[70] Lemke, W., *Term Structure Modeling and Estimation in a State Space Framework*, Springer, 2006.

[71] Litterman, R., Scheinkman, J., "Common factors affecting bond returns", *Journal of Fixed Income*, **1**, 62–74, 1991.

[72] Macdonald, R. & Taylor, M. P., "The Monetary Approach to the Exchange Rate: Rational Expectations, Long Run Equilibrium and Forecasting", *International Monetary Fund Staff papers*, **40,1**,89–102, 1993.

[73] MacDonald R., "Exchange Rate Behavior: Are Fundamental Important?", *The Economic Journal*, **109**, 673–691, 1999.

[74] MacDonald, R. & Marsh, I. W., "On Fundamentals and Exchange Rate: A Casselian Perspective", *Review of Economics and Statistics*, 655–664, 1997.

[75] Mark, N. C., "Exchange Rates and Fundamentals: Evidence of Long-Horizon Predictability", *American Economics Review*, **85(1)**,201–218, 1995.

[76] Merton, R. C., "Theory of Rational Option Pricing", *The Bell Journal of Economics and Management Science*, **4, 1**, 141–183, 1973.

[77] Oksendal, *Stochastic differential equations*, Springer, Berlin, 2005.

[78] Pagan, A. R., Hall, A. D., and Martin, V., "Modeling the term structure", working paper, Bond University, 1995.

[79] Perignon, C., Smith, D. R., and Villa, C., "Why Common Factors in International Bond Returns Are Not So Common", *Journal of International Money and Finance*, **26**, 284–304, 2007.

[80] Piazzesi, M., "Monetary Policy and Macroeconomic Variables in a Modle of the Term Structrue of Interest Rates", working paper, Stanford University, 1998.

[81] Piazzesi, M., "Affine Term Structure Models", working paper, University of Chicago, 2003.

[82] Rapach, D. E. & Wohar, M. E., "Testing the Monetary Model of Exchange Rate Determination: New Evidence from a Century of Data", *Journal of International Economics*, **58, 2**, 359–385, 2002.

[83] Rauch E. J., "Networks versus markets in international trade", *Journal of International Economics*, Vol. **48**, (1999), 7–35.

[84] Schweizer, B., and Wolff, E. F., "On nonparametric Measures of Dependence for Random Variables", *Annals of Statistics*, **9**, 870-885, 1981.

[85] Shriram, Murray, Hyun, and Claffy (2005), "Comparison of Public End-to-End Bandwidth Estimation Tools on High-Speed Links", CAIDA, San Diego Supercomputer Center.

[86] Singh, N., and Boughton, P. D., "Measuring Website Globalization: A Cross-Sectional Country and Industry Level Analysis", *Journal of Website Promotion*, Vol. **1(3)**, 2005, 3–20.

[87] Steenkiste, R. J., and Foresi, S., "Arrow–Debreu Prices for Affine Models", *Working Paper*, April, 1999.

[88] Taylor, M., "The Economics of Exchange Rate", *Journal of Economic Literature*, Vol. 33, Issue 1, 13–47, March, 1995.

[89] Tsay, R. S., *Analysis of Financial Time Series*, 2nd Edition, Wiley, 2005.

[90] Vasicek, O. A., "An equilibrium characteristization of the Term Structure", *J. Financial Economics*, **5**, 177-188, 1977.

[91] Wagner, C. S., and Leydesdorff, L., "Measuring the Globalization of Knowledge Networks", working paper, Sep. 2006.

[92] WARD (1996), "The Effect of the Internet on Political Institutions", *Oxford Journals*, Volume **5**, Number 4 ,Page. 1127-1141.

[93] ZIVOT, E. AND WANG, J., *Modeling Financial Time series with S-Plus*, Springer-Verlag, New York, 2003.